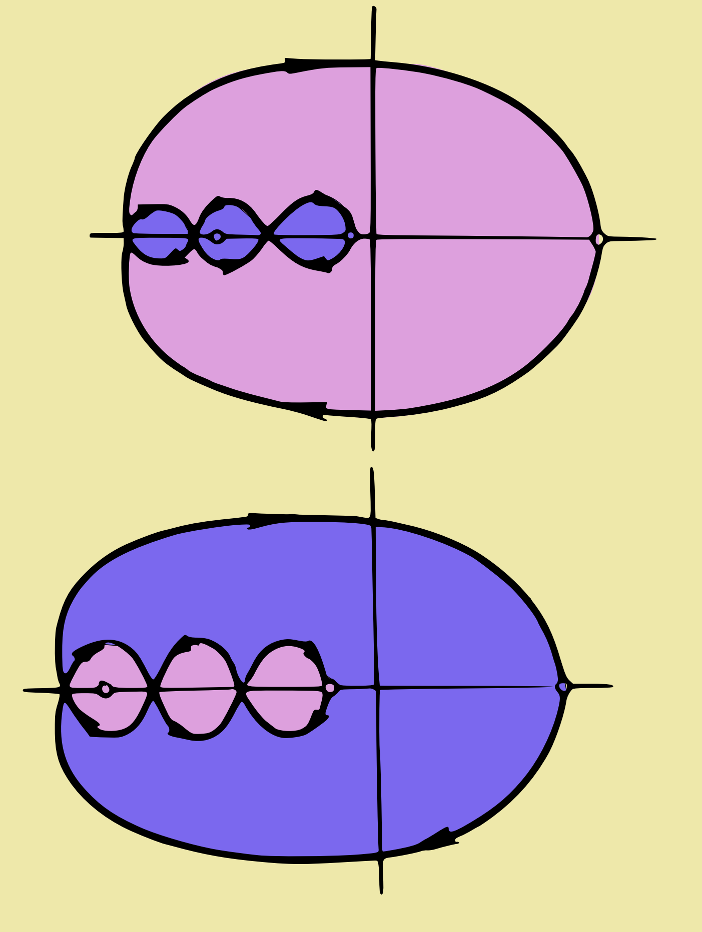
I.N. Roïtenberg

THÉORIE DE LA COMMANDE AUTOMATIQUE



Éditions Mir Moscou

Я. Н. РОЙТЕНБЕРГ

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ

ИЗДАТЕЛЬСТВО « НАУКА »
МОСКВА

I. N. ROÏTENBERG

THÉORIE DU CONTRÔLE AUTOMATIQUE

EDITIONS « MIR »
MOSCOU 1974

Traduit du russe
par VALENTIN POLONSKI

PRÉFACE

La théorie du contrôle automatique est devenue de nos jours une discipline très vaste, ses résultats présentant un grand intérêt théorique et pratique. Actuellement l'étude de cette branche scientifique s'impose non seulement aux spécialistes, mais encore aux étudiants de nombreuses spécialités. L'ouvrage que nous proposons au lecteur est un cours professé par l'auteur aux étudiants en Mécanique de la faculté de Mécanique et des Mathématiques de l'Université de Moscou.

En guidant ses lecteurs depuis les notions initiales de la théorie de la commande jusqu'aux problèmes actuels, l'auteur a tenu également à rendre évidente la liaison qui existe entre les problèmes et les méthodes de la théorie de la commande du mouvement d'une part et les principes généraux de la mécanique analytique de l'autre.

Il nous paraît inutile de donner ici un énoncé sommaire de nombreux problèmes de la théorie de la commande moderne. L'auteur a pris à tâche de traiter avec force détails et le plus rigoureusement possible une série de problèmes actuels de cette théorie. De plus, les problèmes exposés dans ce livre, et qui présentent un intérêt particulier, peuvent servir de base à l'étude ultérieure des monographies et des articles consacrés aux divers aspects particuliers de cette théorie et notamment, aux problèmes de commande des systèmes à paramètres répartis, de la théorie de poursuite, de stabilisation optimale, d'adaptation, etc.

L'auteur tient à exprimer sa profonde reconnaissance à Nikolaï Krassovski de l'Académie des Sciences de l'U.R.S.S. qui a révisé le manuscrit et a fait des remarques précieuses qui ont beaucoup contribué à améliorer le présent ouvrage.

La traduction française de cet ouvrage contient quelques compléments inexistants dans l'édition russe. L'auteur a remanié tout particulièrement le chapitre six relatif à la théorie des systèmes stochastiques. Les paragraphes ajoutés (§§ 27, 29, 31, 32, 33 et 34) portent sur les problèmes de filtrage optimal des processus aléatoires multidimensionnels et les problèmes de commande optimale des systèmes stochastiques.

CHAPITRE PREMIER

SYSTÈMES COMMANDÉS LINÉAIRES

§ 1. Systèmes commandés monovariables

1. Systèmes à un degré de liberté. L'équation du mouvement d'un système à un degré de liberté, dont l'énergie cinétique est $T = \frac{1}{2} k_0 \dot{x}^2$, l'énergie potentielle $V = \frac{1}{2} k_2 x^2$ et la force non conservative supplémentaire $Q = -k_1 \dot{x} + z(t)$, s'écrira

$$k_0 \ddot{x} + k_1 \dot{x} + k_2 x = z(t). \tag{1.1}$$

En désignant par D l'opérateur de différentiation par rapport au temps

$$D = \frac{d}{dt}, \tag{1.2}$$

on peut récrire l'équation (1) sous la forme:

$$(k_0 D^2 + k_1 D + k_2)\, x = z(t) \tag{1.3}$$

ou

$$x = \frac{1}{k_0 D^2 + k_1 D + k_2}\, z(t). \tag{1.4}$$

La fonction

$$\Phi(D) = \frac{1}{k_0 D^2 + k_1 D + k_2} \tag{1.5}$$

est une fonction rationnelle de l'opérateur de différentiation D, qui s'appelle *fonction de transfert* du système.

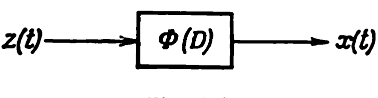


Fig. 1.1

En portant l'expression (5) dans (4), on obtient

$$x = \Phi(D)\, z(t). \tag{1.6}$$

La relation (6) est équivalente à l'équation différentielle initiale (1). Cette relation correspond au diagramme fonctionnel de la figure 1.1.

La fonction $z(t)$ est dite signal *d'entrée*, et la fonction $x(t)$, signal *de sortie* du système.

Cherchons la solution de l'équation différentielle (1). Si

$$k_1^2 - 4k_0k_2 < 0, \tag{1.7}$$

les racines de l'équation caractéristique

$$k_0\lambda^2 + k_1\lambda + k_2 = 0 \tag{1.8}$$

sont

$$\lambda_1, \lambda_2 = -\varepsilon \pm i\omega, \tag{1.9}$$

où

$$\varepsilon = \frac{k_1}{2k_0}, \quad \omega = \sqrt{\frac{k_2}{k_0} - \frac{k_1^2}{4k_0^2}}. \tag{1.10}$$

Sous la condition (7), la solution de (1) s'écrira

$$x(t) = x(t_0)\, e^{-\varepsilon(t-t_0)} \cos\omega(t-t_0) + \frac{1}{\omega}[\dot{x}(t_0) + \varepsilon x(t_0)]\, e^{-\varepsilon(t-t_0)} \sin\omega(t-t_0) + \int_{t_0}^{t} \frac{1}{k_0\omega} e^{-\varepsilon(t-\tau)} \sin\omega(t-\tau)\, z(\tau)\, d\tau. \tag{1.11}$$

Lorsque les conditions initiales sont nulles, c'est-à-dire si

$$x(t_0) = 0, \quad \dot{x}(t_0) = 0, \tag{1.12}$$

la loi du mouvement du système considéré sera

$$x(t) = \int_{t_0}^{t} \frac{1}{k_0\omega} e^{-\varepsilon(t-\tau)} \sin\omega(t-\tau)\, z(\tau)\, d\tau. \tag{1.13}$$

Désignons par $k(t)$ la fonction suivante

$$k(t) = \frac{1}{k_0\omega} e^{-\varepsilon t} \sin\omega t \qquad (t \geqslant 0). \tag{1.14}$$

Cette fonction s'appelle *fonction de poids* du système (1). Il résulte de l'expression (11) que $k(t - t_0)$ est la loi du mouvement du système pour

$$z(t) \equiv 0,$$

tandis que les conditions initiales sont

$$x(t_0) = 0, \quad \dot{x}(t_0) = \frac{1}{k_0}. \tag{1.15}$$

D'après l'expression (11) on peut conclure que $k(t - t_0)$ sera également la loi du mouvement du système dans le cas où

$$x(t_0) = 0, \quad \dot{x}(t_0) = 0, \quad z(t) = \delta(t - t_0), \tag{1.16}$$

où $\delta(t)$ est une fonction impulsion (fonction delta de Dirac). Pour cette raison la fonction $k(t)$ est dite également *fonction de transition impulsionnelle* du système.

Des relations (16) il suit que jusqu'à l'instant $t = t_0$, i.e. jusqu'à l'instant de l'application de l'impulsion unitaire, le système se trouvait en équilibre. Il en résulte que nous devons poser $k(t - t_0) = 0$ pour $t < t_0$, c'est-à-dire que la fonction de poids est nulle lorsque la variable est négative.

Ainsi l'expression (14) qui définit la fonction de poids $k(t)$ doit être complétée par la relation

$$k(t) = 0 \text{ pour } t < 0. \tag{1.17}$$

Conformément à (14) l'expression (13) pourra s'écrire

$$x(t) = \int_{t_0}^{t} k(t-\tau) z(\tau) d\tau. \tag{1.18}$$

L'expression (18) peut être transformée en posant

$$\xi = t - \tau. \tag{1.19}$$

Si l'on retient que t est ici un paramètre, on aura $d\xi = -d\tau$. Puisque $\xi = 0$ pour $\tau = t$ et $\xi = t - t_0$ pour $\tau = t_0$, on obtient

$$x(t) = \int_{0}^{t-t_0} k(\xi) z(t-\xi) d\xi. \tag{1.20}$$

De cette façon le système où

$$x(t_0) = 0, \quad \dot{x}(t_0) = 0,$$

a pour loi du mouvement

$$x(t) = \int_{t_0}^{t} k(t-\tau) z(\tau) d\tau = \int_{0}^{t-t_0} k(\xi) z(t-\xi) d\xi. \tag{1.21}$$

Considérons maintenant un cas limite. Soit

$$t_0 = -\infty,$$

ce qui traduit le fait que l'application du signal d'entrée a eu lieu il y a un temps infini. Si $\varepsilon > 0$, c'est-à-dire si les oscillations libres du système (1) s'amortissent asymptotiquement, l'expression (11) devient

$$x(t) = \lim_{t_0 \to -\infty} \int_{t_0}^{t} k(t-\tau) z(\tau) d\tau = \lim_{t_0 \to -\infty} \int_{0}^{t-t_0} k(\xi) z(t-\xi) d\xi,$$

ce qui pour raison de simplicité s'écrira sous la forme abrégée

$$x(t) = \int_{-\infty}^{t} k(t-\tau) z(\tau) d\tau = \int_{0}^{\infty} k(\xi) z(t-\xi) d\xi. \qquad (1.22)$$

L'expression (22) traduit le processus *stationnaire* du système.

Exemple. Trouvons à titre d'exemple le processus stationnaire du système décrit par l'équation différentielle (1)

$$k_0\ddot{x} + k_1\dot{x} + k_2 x = z(t)$$

pour le cas où

$$z(t) \equiv k_2 m = \text{const.}$$

D'après (22) et (14)

$$\lim_{t_0 \to -\infty} x(t) = \lim_{t_0 \to -\infty} \int_{0}^{t-t_0} k(\xi) z(t-\xi) d\xi$$

ou

$$\lim_{t_0 \to -\infty} x(t) = \frac{k_2 m}{k_0 \omega} \lim_{t_0 \to -\infty} \int_{0}^{t-t_0} e^{-\varepsilon\xi} \sin \omega\xi \, d\xi =$$

$$= \frac{k_2 m}{k_0 \omega} \lim_{t_0 \to -\infty} \left[\frac{-\varepsilon \sin \omega\xi - \omega \cos \omega\xi}{\varepsilon^2 + \omega^2} e^{-\varepsilon\xi} \right]_0^{t-t_0}.$$

Etant donné que d'après (10) $\varepsilon > 0$, $\varepsilon^2 + \omega^2 = \frac{k_2}{k_0}$, l'expression obtenue aura pour la forme

$$\lim_{t_0 \to -\infty} x(t) = m.$$

2. Fonction de poids et fonction de transfert. Un système décrit par les équations différentielles linéaires à coefficients constants est dit système stationnaire. Pour un tel système la transformation de Laplace permet d'établir une correspondance assez intéressante entre la fonction de transfert et la fonction de poids.

Rappelons [90] qu'on appelle transformée de Laplace de la fonction $x(t)$ d'une variable réelle t la fonction d'une variable complexe p définie par la relation

$$\mathcal{L}[x(t)] = \int_{0}^{\infty} x(t) e^{-pt} dt. \qquad (1.23)$$

Dans le calcul opérationnel de Carson-Heaviside, la transformée $\xi(p)$ de la fonction $x(t)$ est définie par la relation

$$\frac{\xi(p)}{p} = \int_{0}^{\infty} x(t) e^{-pt} dt, \qquad (1.24)$$

ce qui, en principe, ne diffère en rien de la transformation de Laplace mais présente un avantage tel que la fonction $x(t) \equiv c$, où c = const, a pour transformée la même constante: $\xi(p) \equiv c$. La conformité opérationnelle entre les fonctions $\xi(p)$ et $x(t)$ (la fonction $x(t)$ s'appelle original) se note

$$\xi(p) \doteq x(t) \quad \text{ou} \quad x(t) \doteq \xi(p). \tag{1.25}$$

Si l'on connaît la fonction $\xi(p)$, l'original qui lui correspond est déterminé par la formule de Riemann-Mellin

$$x(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\xi(p)}{p} e^{pt}\, dp \qquad (t \geqslant 0), \tag{1.26}$$

où l'intégration s'opère dans le plan de la variable complexe p le long de la droite parallèle à l'axe imaginaire et distante de ce dernier de c. Cette droite se situe à droite de tous les points singuliers de la fonction $\xi(p)/p$.

Considérons maintenant pour le système (1) étudié ci-dessus la fonction $\Phi(p)$ de la variable complexe p, ladite fonction étant obtenue à partir de la fonction de transfert $\Phi(D)$ si l'on y remplace D par p. D'après (5) la fonction $\Phi(p)$ sera de la forme:

$$\Phi(p) = \frac{1}{k_0 p^2 + k_1 p + k_2}. \tag{1.27}$$

En tenant compte des relations (7), (8) et (9), l'expression (27) peut s'écrire

$$\Phi(p) = \frac{1}{k_0 (p+\varepsilon - i\omega)(p+\varepsilon+i\omega)} = \frac{1}{k_0 [(p+\varepsilon)^2+\omega^2]}. \tag{1.28}$$

D'après (28)

$$p\Phi(p) = \frac{1}{k_0\omega} \frac{p\omega}{(p+\varepsilon)^2+\omega^2}. \tag{1.29}$$

Ainsi que l'on connaît [17],

$$\frac{p\omega}{(p+\varepsilon)^2+\omega^2} \doteq e^{-\varepsilon t} \sin \omega t.$$

De cette façon,

$$p\Phi(p) \doteq \frac{1}{k_0\omega} e^{-\varepsilon t} \sin \omega t, \tag{1.30}$$

par conséquent, en vertu de (14), on aboutit à la relation

$$p\Phi(p) \doteq k(t). \tag{1.31}$$

C'est précisément la relation entre la fonction de transfert $\Phi(D)$ et la fonction de poids $k(t)$ du système (1) qui nous intéresse.

Nous allons montrer dans ce qui suit que la relation (31) a lieu pour tout système linéaire stationnaire à nombre fini de degrés de liberté.

La fonction de transfert $\Phi(D)$ ou la fonction de poids $k(t)$ qui lui correspond représentent la caractéristique dynamique principale d'un tel système. Ces fonctions caractérisent aussi bien les oscillations propres du système que sa capacité de reproduire les signaux extérieurs. Les fonctions $\Phi(D)$ et $k(t)$ sont absolument équivalentes.

3. Lieu de transfert. Notons encore qu'en vertu de la formule (26) de Riemann-Mellin

$$k(t)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\Phi(p)\,e^{pt}\,dp \quad (t\geqslant 0). \tag{1.32}$$

Comme le montrent (28) et (11), dans un système où tous les pôles de la fonction $\Phi(p)$ sont situés à gauche de l'axe imaginaire, les oscillations libres s'amortissent asymptotiquement. En calculant pour un tel système l'intégrale (32) on peut poser $c = 0$.

Les relations (31) et (24) entraînent que si l'on connaît la fonction de poids $k(t)$ du système la fonction $\Phi(p)$ peut être définie par la formule

$$\Phi(p)=\int_{0}^{\infty}k(t)\,e^{-pt}\,dt. \tag{1.33}$$

En posant $p = i\omega$ et en tenant compte que $k(t) = 0$ pour $t < 0$, dans le cas d'un système tel que les pôles de la fonction $\Phi(p)$ sont situés à gauche de l'axe imaginaire, on peut écrire la transformée de Fourier pour la fonction de poids $k(t)$ sous la forme

$$\Phi(i\omega)=\int_{-\infty}^{\infty}k(t)\,e^{-i\omega t}\,dt. \tag{1.34}$$

La fonction $\Phi(i\omega)$ s'appelle *lieu de transfert* ou *caractéristique fréquentielle* du système.

Dans un système pour lequel les pôles de la fonction $\Phi(p)$ ne se trouvent pas tous à gauche de l'axe imaginaire, l'amortissement asymptotique des oscillations libres n'a pas lieu et la formule (34) ne peut pas servir pour déterminer le lieu de transfert $\Phi(i\omega)$ du fait que pour un tel système l'intégrale impropre

$$\int_{-\infty}^{\infty}|k(t)|\,dt$$

est divergente.

4. Système commandé en boucle fermée. La figure 1.2 représente le diagramme fonctionnel d'un système asservi dont le signal de sortie est x. Le signal d'entrée du système est une fonction du temps

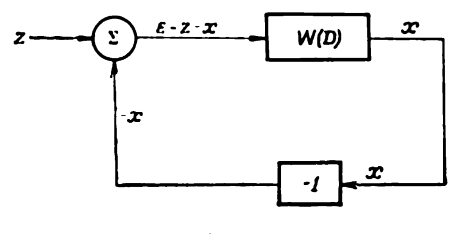


Fig. 1.2

$z(t)$ inconnue à l'avance. Le système est destiné à maintenir à tout instant t la valeur $x(t)$ du signal de sortie à un niveau assez proche de celui du signal d'entrée $z(t)$, en rendant ainsi suffisamment petit l'écart

$$\varepsilon = z - x$$

du système asservi.

La construction du système asservi, conforme au diagramme fonctionnel de la figure 1.2, peut être différente. Par exemple (fig. 1.3), on peut concevoir un arbre posé sur des paliers (dit arbre

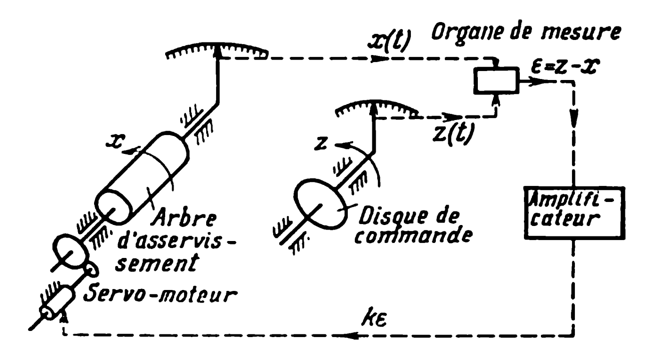


Fig. 1.3

d'asservissement), dont le centre de gravité se situe sur l'axe de rotation. L'arbre d'asservissement est entraîné en rotation par un moteur électrique à courant continu et à excitation séparée (servomoteur). La tension appliquée au circuit de l'induit est proportionnelle à l'écart de l'asservissement. Dans notre exemple, x est l'angle de rotation de l'arbre d'asservissement, et $z(t)$, l'angle de rotation d'un certain disque de commande dont le mouvement doit être reproduit par l'arbre d'asservissement.

Le diagramme fonctionnel de la figure 1.2 représente un système commandé *en boucle fermée*. Les équations du mouvement du système asservi envisagé peuvent s'écrire

$$\left.\begin{aligned}(k_0D^2 + k_1D)\, x &= k_2\varepsilon, \\ \varepsilon &= z - x,\end{aligned}\right\} \tag{1.35}$$

où $z = z\,(t)$ est le signal d'entrée du système. La première équation (35) est celle du mouvement de l'arbre d'asservissement. La deuxième équation (35) définit la valeur de l'écart. En introduisant les notations

$$W\,(D) = \frac{k_2}{k_0D^2 + k_1D}, \tag{1.36}$$

on peut représenter le système (35) sous la forme

$$x = W\,(D)\,(z - x)$$

ou

$$[1 + W\,(D)]\, x = W\,(D)\, z. \tag{1.37}$$

Avec les notations

$$\Phi\,(D) = \frac{W\,(D)}{1 + W\,(D)}, \tag{1.38}$$

l'équation (37) deviendra

$$x = \Phi\,(D)\, z\,(t). \tag{1.39}$$

L'équation (39) s'appelle équation du système en boucle fermée. La fonction $\Phi\,(D)$ est dite *fonction de transfert du système*

$z(t)$ → $\Phi(D)$ → $x(t)$

Fig. 1.4

en boucle fermée. Le diagramme de la figure 1.2 est ainsi ramené à la forme (fig. 1.4) étudiée au point 1.

Pour le système asservi considéré, la fonction $\Phi\,(D)$ d'après (36) et (38) a pour expression

$$\Phi\,(D) = \frac{k_2}{k_0D^2 + k_1D + k_2}, \tag{1.40}$$

et l'équation (39) est équivalente à l'équation différentielle

$$(k_0D^2 + k_1D + k_2)\, x = k_2\, z\,(t). \tag{1.41}$$

5. Système commandé en boucle ouverte. Supposons maintenant que dans un système asservi la chaîne de réaction soit ouverte (fig. 1.5). La tension appliquée au circuit de l'induit est cette fois proportionnelle non pas à l'écart mais au signal d'entrée $z\,(t)$. Le dia-

gramme fonctionnel du système asservi de ce type, dit système commandé en boucle ouverte, est donné par la figure 1.5.

D'après (35) les équations du mouvement du système asservi auront dans ce cas pour expression

$$(k_0D^2 + k_1D)\,x = k_2\,z\,(t). \quad (1.42)$$

D'après (36) l'équation (42) peut être ramenée à la forme

$$x = W(D)\,z\,(t). \quad (1.43)$$

L'équation (43) est dite équation d'un système en boucle ouverte, la fonction $W\,(D)$ étant sa *fonction de transfert*.

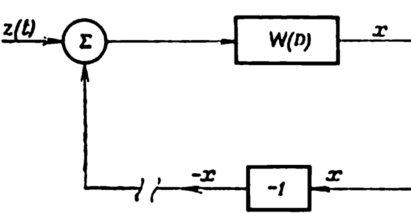


Fig. 1.5

L'expression (38) obtenue ci-dessus

$$\Phi'(D) = \frac{W\,(D)}{1 + W\,(D)}$$

établit la relation entre les fonctions de transfert des systèmes en boucle fermée et en boucle ouverte.

6. Reproduction du signal d'entrée converti. Si le système de commande consiste à reproduire non pas le signal d'entrée $z\,(t)$

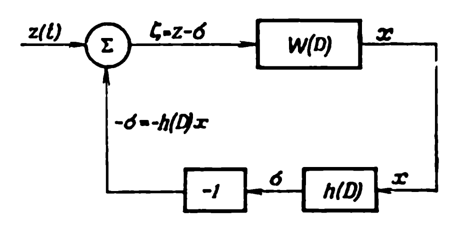


Fig. 1.6

lui-même mais son signal converti $H\,(D)z(t)$, $H\,(D)$ étant une fonction rationnelle de l'opérateur de différentiation D, on peut le réaliser en insérant dans la boucle de retour un dispositif dont la fonction de transfert est $h\,(D)$, où

$$h\,(D) = \frac{1}{H\,(D)}. \quad (1.44)$$

On obtient ainsi le diagramme de la figure 1.6. Le signal de l'écart est

$$\zeta = z - h\,(D)\,x. \quad (1.45)$$

Les équations du système en boucle fermée seront

$$\left.\begin{aligned} (k_0D^2 + k_1D)\,x &= k_2\zeta, \\ \zeta &= z - h\,(D)\,x. \end{aligned}\right\} \quad (1.46)$$

D'après la relation (36), nous récrivons ces équations sous la forme

$$x = W(D)[z - h(D)x]$$

ou

$$[1 + W(D)h(D)]x = W(D)z(t). \tag{1.47}$$

La fonction de transfert du système en boucle fermée s'écrira alors

$$\Psi(D) = \frac{W(D)}{1 + W(D)h(D)}. \tag{1.48}$$

En vertu de (48), l'équation (47) devient:

$$x = \Psi(D)z(t). \tag{1.49}$$

Le système de commande en boucle ouverte est représenté sur la

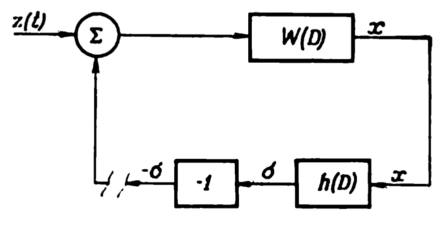


Fig. 1.7

figure 1.7. Les équations de ce système s'écriront

$$\left.\begin{aligned} x &= W(D)z(t), \\ \sigma &= h(D)x, \end{aligned}\right\} \tag{1.50}$$

d'où

$$\sigma = h(D)W(D)z(t). \tag{1.51}$$

De cette façon la fonction de transfert du système en boucle ouverte est de la forme

$$S(D) = h(D)W(D). \tag{1.52}$$

7. Système monovariable à nombre fini de degrés de liberté. Parmi les systèmes de commande, le système à un degré de liberté examiné précédemment est le plus simple. De nombreux systèmes possèdent plusieurs degrés de liberté. Un système qui a pour tâche d'assurer que la loi de variation dans le temps seulement d'une seule des coordonnées $q_j = q_j(t)$ de ce système soit proche d'une certaine fonction du temps $z(t)$ (en général, inconnue à l'avance), la loi de variation des autres coordonnées généralisées $q_i(t)$ $(i = 1, 2, \ldots; i \neq j)$ n'étant pas soumise à ces contraintes, est dit *monovariable* (*à une variable*).

Son mouvement est gouverné par les équations suivantes

$$\dot{q}_s = \sum_{k=1}^{n} A_{sk} q_k + m_s [z(t) - q_j] \quad (s = 1, \ldots, n). \tag{1.53}$$

Le système des équations différentielles scalaires (53) est équivalent à l'équation différentielle vectorielle

$$\dot{q} = Aq + m [z(t) - q_j], \tag{1.54}$$

où q, A et m désignent les matrices

$$q = \begin{bmatrix} q_1 \\ \cdots \\ q_n \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \ldots & A_{nn} \end{bmatrix}, \quad m = \begin{bmatrix} m_1 \\ \cdots \\ m_n \end{bmatrix}. \tag{1.55}$$

L'équation (54) peut être mise sous la forme

$$(ED - A) q = m [z(t) - q_j] \quad \left(D = \frac{d}{dt}\right), \tag{1.56}$$

où E est la matrice unité. Il en résulte

$$q = \frac{B(D) m [z(t) - q_j]}{f(D)}, \tag{1.57}$$

où $B(D)$ est la matrice adjointe de la matrice $ED - A$, et $f(D)$, le déterminant de cette dernière:

$$f(D) = \det (ED - A). \tag{1.58}$$

De (57) il ressort que

$$q_j = \frac{\sum_{l=1}^{n} B_{jl}(D) m_l [z(t) - q_j]}{f(D)}. \tag{1.59}$$

De cette façon l'équation différentielle vérifiée par la coordonnée généralisée q_j qui nous intéresse, s'écrira

$$f(D) q_j = g(D) [z(t) - q_j], \tag{1.60}$$

où

$$g(D) = \sum_{l=1}^{n} B_{jl}(D) m_l. \tag{1.61}$$

Etant donné que $B_{jl}(D)$ est un cofacteur de l'élément situé à l'intersection de la l-ième ligne et la j-ième colonne du déterminant de la matrice $ED - A$, le degré du polynôme $g(D)$ est inférieur à celui du polynôme $f(D)$. Les polynômes $f(D)$ et $g(D)$ peuvent être mis sous la forme

$$\left.\begin{aligned} f(D) &= a_0 D^n + a_1 D^{n-1} + \ldots + a_{n-1} D + a_n, \\ g(D) &= b_0 D^\nu + b_1 D^{\nu-1} + \ldots + b_{\nu-1} D + b_\nu \quad (\nu < n) \end{aligned}\right\}. \tag{1.62}$$

Si le système est stationnaire, les coefficients des polynômes (62) a_i $(i = 0, 1, \ldots, n)$, b_k $(k = 0, 1, \ldots, \nu)$ sont constants.

D'une façon analogue à ce qui a été exposé ci-dessus, dans le système envisagé, q_j peut être considéré comme signal de sortie et $z(t)$, comme signal d'entrée.

Pour rendre plus commode la comparaison avec les systèmes examinés, remplaçons q_j par x et écrivons l'équation (60) sous la forme suivante:

$$\left.\begin{aligned} f(D)\,x &= g(D)\,\varepsilon, \\ \varepsilon &= z - x, \end{aligned}\right\} \qquad (1.63)$$

où ε désigne l'écart. Si l'on introduit les notations

$$G(D) = \frac{g(D)}{f(D)} = \frac{b_0 D^{\nu} + b_1 D^{\nu-1} + \ldots + b_{\nu-1} D + b_{\nu}}{a_0 D^n + a_1 D^{n-1} + \ldots + a_{n-1} D + a_n}, \qquad (1.64)$$

on peut ramener le système des équations (63) à la forme

$$x = G(D)\,(z - x)$$

ou

$$[1 + G(D)]\,x = G(D)\,z(t). \qquad (1.65)$$

Si l'on note

$$L(D) = \frac{G(D)}{1 + G(D)}, \qquad (1.66)$$

l'équation (65) devient

$$x = L(D)\,z(t). \qquad (1.67)$$

L'équation (67) est celle d'un système en boucle fermée monovariable à nombre fini de degrés de liberté (fig. 1.8). La fonction $L(D)$ est la *fonction de transfert du système en boucle fermée.*

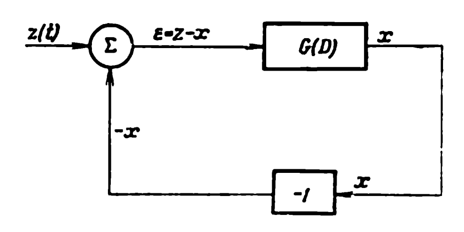


Fig. 1.8

Puisque d'après (64) et (66)

$$L(D) = \frac{g(D)}{f(D) + g(D)},$$

l'équation (67) est équivalente à l'équation différentielle suivante

$$[f(D) + g(D)]\,x = g(D)\,z(t). \qquad (1.68)$$

Le diagramme fonctionnel du système *en boucle ouverte* est dans ce cas de la forme représentée sur la figure 1.9. Son mouvement est décrit par l'équation

$$f(D)\,x = g(D)\,z(t). \tag{1.69}$$

Si l'on tient compte de l'expression (64), on peut ramener l'équation (69) à la forme

$$x = G(D)\,z(t). \tag{1.70}$$

L'équation (70) est l'équation du système commandé en boucle ouverte et la fonction $G(D)$ est sa *fonction de transfert.*

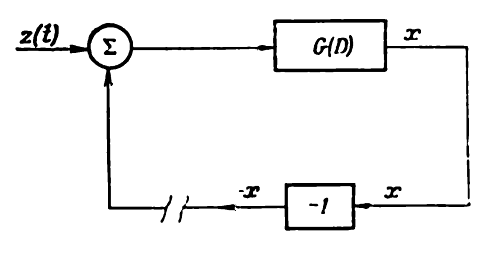


Fig. 1.9

8. Une remarque sur l'intégration des équations du mouvement d'un système monovariable. Dans ce qui précède nous avons passé du système d'équations différentielles initial (53) à l'équation (60). La fonction $z(t)$ fait partie de l'équation (60) sous le signe de l'opérateur différentiel, bien que le système d'équations (53) ne comportait pas les dérivées de $z(t)$. La solution de l'équation différentielle (60) dépendra de la valeur initiale de la fonction $z(t)$ elle-même et des valeurs initiales des dérivées de $z(t)$ (jusqu'à la valeur initiale de la dérivée de $z(t)$ dont le degré est d'un ordre inférieur à celui de l'opérateur différentiel $g(D)$). Cependant les valeurs initiales de la fonction $z(t)$ elle-même et de ses dérivées (si ces dérivées n'entrent pas dans le système initial) n'interviennent pas dans la solution du système d'équations initial.

Pour rendre clair le sens de cette question, examinons l'exemple suivant. Soit le système d'équations

$$\left.\begin{aligned} \dot{q}_1 + kq_2 &= 0, \\ \dot{q}_2 - kq_1 &= z(t). \end{aligned}\right\} \tag{1.71}$$

En éliminant des équations (71) la variable q_1 on obtient l'équation différentielle du deuxième ordre

$$\ddot{q}_2 + k^2 q_2 = \dot{z}(t). \tag{1.72}$$

Intégrons l'équation (72) en appliquant les méthodes du calcul opérationnel. Introduisons les notations

$$\eta_2(p) \doteqdot q_2(t), \quad \zeta(p) \doteqdot z(t). \tag{1.73}$$

Etant donné que

$$\left.\begin{aligned} \ddot{q}_2(t) &\doteqdot p^2\eta_2(p) - p^2 q_2(0) - p\dot{q}_2(0), \\ \dot{z}(t) &\doteqdot p\zeta(p) - pz(0), \end{aligned}\right\} \tag{1.74}$$

à l'équation différentielle (72) correspondra la transformation

$$(p^2 + k^2)\,\eta_2(p) = p\zeta(p) + p^2 q_2(0) + p\dot{q}_2(0) - pz(0). \tag{1.75}$$

D'où

$$\eta_2(p) = \frac{1}{p}\,\frac{p^2}{p^2+k^2}\,\zeta(p) + q_2(0)\,\frac{p^2}{p^2+k^2} + \frac{1}{k}\,[\dot{q}_2(0) - z(0)]\,\frac{pk}{p^2+k^2}. \tag{1.76}$$

Si l'on retient que

$$\frac{p^2}{p^2+k^2} \doteqdot \cos kt, \quad \frac{pk}{p^2+k^2} \rightarrow \sin kt$$

et que d'après le théorème de multiplication des transformées

$$\frac{1}{p}\,\frac{p^2}{p^2+k^2}\,\zeta(p) \doteqdot \int_0^t \cos k(t-\tau)\, z(\tau)\, d\tau,$$

on obtiendra, en vertu de (76), la loi suivante de la variation dans le temps de la coordonnée q_2:

$$q_2(t) = \frac{1}{k}\,[\dot{q}_2(0) - z(0)]\sin kt + q_2(0)\cos kt + \int_0^t \cos k(t-\tau)\, z(\tau)\, d\tau. \tag{1.77}$$

Notons maintenant que le système d'équations initial (71) entraîne

$$\frac{1}{k}\,[\dot{q}_2(0) - z(0)] = q_1(0), \tag{1.78}$$

et, par conséquent, la solution de (77) pourra être mise sous la forme

$$q_2(t) = q_1(0)\sin kt + q_2(0)\cos kt + \int_0^t \cos k(t-\tau)\, z(\tau)\, d\tau. \tag{1.79}$$

La valeur initiale $z(0)$ de la fonction $z(t)$ n'intervient pas dans l'expression (79).

Ainsi, pour $q_1(0) = 0$, $q_2(0) = 0$ les deux premiers termes de l'expression (77) qui correspondent aux oscillations libres du système, s'éliminent et (77) devient

$$q_2(t) = \int_0^t \cos k(t-\tau)\, z(\tau)\, d\tau. \tag{1.80}$$

§ 2. Systèmes commandés multivariables

1. Système commandé en boucle fermée. Un système commandé dont le signal d'entrée et le signal de sortie sont des vecteurs est dit *multivariable* ou *à plusieurs variables* (fig. 2.1).

Son diagramme fonctionnel est représenté sur la figure 2.2. Ici α_1 est un organe comprenant l'objet commandé, les amplificateurs et les éléments d'exécution. Par α_2 on désigne l'organe comprenant l'ensemble des éléments sensibles, c'est-à-dire des appareils qui

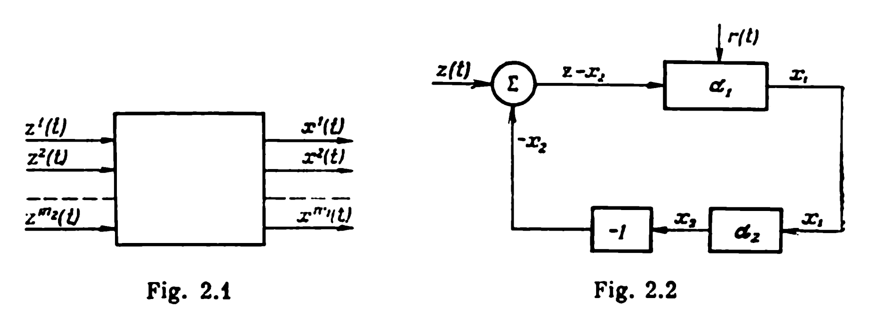


Fig. 2.1 Fig. 2.2

mesurent les paramètres de l'état du système commandé. x_1 est le signal de sortie de l'organe α_1; c'est un vecteur de dimension m_1. Par x_2 est désigné le signal de sortie de l'organe α_2, qui est un vecteur de dimension m_2.

Les équations différentielles qui décrivent les processus se déroulant dans les organes α_1 et α_2 (en suivant le point de vue général de Lagrange, nommons-les équations du mouvement du système) peuvent être mises sous la forme du système d'équations différentielles vectorielles suivant

$$\left.\begin{aligned} f_1(D)\,x_1 &= e_1(D)\,(z - x_2) + e_r(D)\,r, \\ f_2(D)\,x_2 &= e_2(D)\,x_1. \end{aligned}\right\} \qquad (2.1)$$

Ici $f_1(D)$ est une matrice $m_1 \times m_1$ dont les éléments sont des polynômes en D (où $D = d/dt$) à coefficients constants. D'une façon analogue, $f_2(D)$ est une matrice polynomiale $m_2 \times m_2$; $e_1(D)$, une matrice $m_1 \times m_2$; $e_2(D)$, une matrice $m_2 \times m_1$.

L'entrée $z(t)$ est un vecteur de dimension m_2; $r(t)$ désigne le vecteur de dimension m_r qui représente l'ensemble de bruits (perturbations) appliqués au système. $e_r(D)$ est une matrice d'ordre $m_1 \times m_r$.

Le vecteur $z - x_2$ est l'écart du système. Le système consiste à assurer à chaque instant un écart suffisamment petit. Les critères permettant d'apprécier la qualité de la réalisation de cette tâche

diffèrent suivant le système commandé concret. Dans ce qui suit nous donnons la description de certains de ces critères.

2. Déterminant caractéristique d'un système en boucle fermée. L'entrée z et la perturbation r sont certaines (inconnues à l'avance) fonctions du temps

$$z = z(t), \quad r = r(t).$$

Lorsqu'on étudie la stabilité d'un système, on ne considère que ses mouvements propres. En posant $z(t) \equiv 0$, $r(t) \equiv 0$, on obtient à partir de (1) les équations des oscillations libres du système

$$\left.\begin{aligned} f_1(D)\,x_1 + e_1(D)\,x_2 &= 0, \\ f_2(D)\,x_2 - e_2(D)\,x_1 &= 0. \end{aligned}\right\} \tag{2.2}$$

La matrice opérationnelle du système d'équations (2 s'écrit

$$f(D) = \begin{bmatrix} f_1(D) & e_1(D) \\ -e_2(D) & f_2(D) \end{bmatrix}. \tag{2.3}$$

La matrice $f(D)$ est une matrice partitionnée. Le déterminant de $f(D)$ s'obtient après les transformations élémentaires suivantes. Multiplions la deuxième ligne de la matrice $f(D)$ à gauche par la matrice $-e_1(D)\,f_2^{-1}(D)$ (où $f_2^{-1}(D)$ est l'inverse de $f_2(D)$) et additionnons la ligne obtenue avec la première ligne. La matrice triangulaire $f^*(D)$ ainsi obtenue

$$f^*(D) = \begin{bmatrix} f_1(D) + e_1(D)\,f_2^{-1}(D)\,e_2(D) & 0 \\ -e_2(D) & f_2(D) \end{bmatrix} \tag{2.4}$$

est équivalente à la matrice $f(D)$. Comme les déterminants des matrices équivalentes ne diffèrent que par le coefficient constant, l'opérateur $\Delta(D)$ de la matrice opérationnelle $f(D)$ sera donc, à un coefficient constant près,

$$\Delta(D) = \det\,[f_1(D) + e_1(D)\,f_2^{-1}(D)\,e_2(D)]\,\det f_2(D). \tag{2.5}$$

En transformant l'expression entre crochets de (5), on a

$$\begin{aligned} f_1(D) + e_1(D)\,f_2^{-1}(D)\,e_2(D) &= f_1(D)\,[E_{m_1} + f_1^{-1}(D)\,e_1(D)\,f_2^{-1}(D)\,e_2(D)] = \\ &= f_1(D)\,[E_{m_1} + W_1(D)\,W_2(D)] = f_1(D)\,[E_{m_1} + W(D)], \end{aligned} \tag{2.6}$$

où

$$W_i(D) = f_i^{-1}(D)\,e_i(D) = \frac{F_i(D)\,e_i(D)}{\Delta_i(D)}, \quad (i = 1, 2), \tag{2.7}$$

$$W(D) = W_1(D)\,W_2(D) = \frac{F_1(D)\,e_1(D)\,F_2(D)\,e_2(D)}{\Delta_1(D)\,\Delta_2(D)}. \tag{2.8}$$

E_{m_1} désigne ici la matrice unité $m_1 \times m_1$. $F_i(D)$ est la matrice adjointe de $f_i(D)$ et $\Delta_i(D)$, le déterminant de la matrice $f_i(D)$:

$$\Delta_i(D) = \det f_i(D) \quad (i = 1, 2).$$

Les expressions (7) montrent que $W_1(D)$ est une matrice $m_1 \times m_2$, $W_2(D)$, une matrice $m_2 \times m_1$, et $W(D)$, une matrice carrée $m_1 \times m_1$.

La matrice $W_i(D)$ $(i = 1, 2)$ est une *matrice de transfert* de l'organe α_i.

Ainsi, l'expression (5) devient

$$\Delta(D) = \det\{f_1(D)[E_{m_1} + W(D)]\} \det f_2(D). \qquad (2.9)$$

Le déterminant du produit de deux matrices étant égal au produit de leurs déterminants, et, de plus, la multiplication des déterminants étant commutative, l'expression (9) pourra s'écrire ainsi:

$$\Delta(D) = \Delta_1(D)\,\Delta_2(D) \det[E_{m_1} + W(D)]. \qquad (2.10)$$

Le déterminant (10) est le *déterminant caractéristique* du système en boucle fermée.

L'équation caractéristique d'un tel système est la suivante:

$$\Delta(p) = 0. \qquad (2.11)$$

Le système de commande en boucle fermée sera asymptotiquement stable, si toutes les racines de l'équation caractéristique (11), i. e. tous les zéros de la fonction $\Delta(p)$, se trouvent dans le plan de la variable complexe p à gauche de l'axe imaginaire.

3. Equation du contrôle automatique. Reprenons les équations (1)

$$f_1(D)\,x_1 + e_1(D)\,x_2 = e_1(D)\,z + e_r(D)\,r,$$
$$-e_2(D)\,x_1 + f_2(D)\,x_2 = 0$$

et éliminons de ces équations le vecteur x_2. A cet effet, multiplions tous les termes de la deuxième équation (1) à gauche par la matrice $-e_1(D)\,f_2^{-1}(D)$ et ajoutons l'expression ainsi obtenue à la première équation (1). On a

$$[f_1(D) + e_1(D)\,f_2^{-1}(D)\,e_2(D)]\,x_1 = e_1(D)\,z + e_r(D)\,r, \qquad (2.12)$$

ou d'après (6)

$$f_1(D)[E_{m_1} + W(D)]\,x_1 = e_1(D)\,z + e_r(D)\,r. \qquad (2.13)$$

En multipliant le premier et le second membre de (13) à gauche par la matrice $f_1^{-1}(D)$, on obtient

$$[E_{m_1} + W(D)]\,x_1 = f_1^{-1}(D)\,e_1(D)\,z + f_1^{-1}(D)\,e_r(D)\,r. \qquad (2.14)$$

Introduisons les notations

$$f_1^{-1}(D)\,e_r(D) = \frac{F_1(D)\,e_r(D)}{\Delta_1(D)} = R(D). \qquad (2.15)$$

La matrice $R(D)$ sera du type $m_1 \times m_r$. On peut la nommer *matrice de transfert des perturbations.*

Puisque d'après (7)

$$f_1^{-1}(D)\,e_1(D) = \frac{F_1(D)\,e_1(D)}{\Delta_1(D)} = W_1(D),$$

l'équation (14) se met sous la forme

$$[E_{m_1} + W(D)]\, x_1 = W_1(D)\, z + R(D)\, r \tag{2.16}$$

ou

$$x_1 = [E_{m_1} + W(D)]^{-1} W_1(D)\, z + [E_{m_1} + W(D)]^{-1} R(D)\, r. \tag{2.17}$$

L'équation (17) décrit les processus qui ont lieu dans le système en boucle fermée et s'appelle *équation du contrôle automatique*.

4. **Système commandé en boucle ouverte.** La figure 2.3 représente un système commandé en boucle ouverte. Des équations (1)

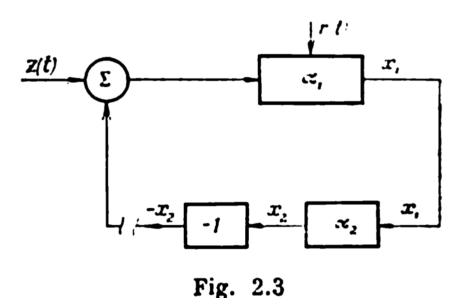


Fig. 2.3

il ressort que les processus qui se déroulent dans un tel système sont décrits par les équations différentielles vectorielles suivantes

$$\left.\begin{aligned} f_1(D)\, x_1 &= e_1(D)\, z(t) + e_r(D)\, r, \\ f_2(D)\, x_2 &= e_2(D)\, x_1. \end{aligned}\right\} \tag{2.18}$$

Les équations des oscillations propres du système en boucle ouverte s'obtiennent à partir des équations (18) en posant $z(t) \equiv 0$, $r(t) \equiv 0$.

Ainsi on a

$$\left.\begin{aligned} f_1(D)\, x_1 &= 0, \\ f_2(D)\, x_2 &= e_2(D)\, x_1. \end{aligned}\right\} \tag{2.19}$$

La matrice opérationnelle du système d'équations (19) s'écrit

$$f_0(D) = \begin{bmatrix} f_1(D) & 0 \\ -e_2(D) & f_2(D) \end{bmatrix}. \tag{2.20}$$

Le déterminant $\Delta_0(D)$ de la matrice $f_0(D)$ est de la forme

$$\Delta_0(D) = \Delta_1(D)\, \Delta_2(D), \tag{2.21}$$

où, de même que précédemment, par $\Delta_i(D)$ $(i = 1, 2)$ est désigné le déterminant de la matrice $f_i(D)$. $\Delta_0(D)$ est le *déterminant caractéristique du système en boucle ouverte.*

Reprenons le système d'équations (18). La première équation est intégrée indépendamment de la deuxième. On peut l'écrire sous la forme

$$x_1 = f_1^{-1}(D)\, e_1(D)\, z + f_1^{-1}(D)\, e_r(D)\, r. \qquad (2.22)$$

En tenant compte des expressions (7) et (15), ramenons (2.22) à la forme

$$x_1 = W_1(D)\, z + R(D)\, r. \qquad (2.23)$$

L'équation (23) décrit le mouvement du système en boucle ouverte à la sortie de l'organe α_2.

L'équation du contrôle automatique (17) diffère de l'expression (23) par la présence de l'opérateur $[E_{m_1} + W(D)]^{-1}$ dans le second membre. Un système commandé doit reproduire le plus exactement possible le signal d'entrée $z(t)$ et être peu sensible aux perturbations $r(t)$. L'opérateur $W(D)$ doit être construit de façon à satisfaire le mieux possible à ces contraintes.

5. Interprétation des opérateurs matriciels $W(D)$ et $E_{m_1} + W(D)$. Considérons un système de commande dans lequel la liaison directe entre les éléments α_1 et α_2 est coupée. Soit

$$z(t) \equiv 0, \quad r(t) \equiv 0, \qquad (2.24)$$

et soit $\vartheta(t)$ un certain signal, représentant un vecteur de dimension m_1 (fig. 2.4), appliqué à l'entrée de l'organe α_2.

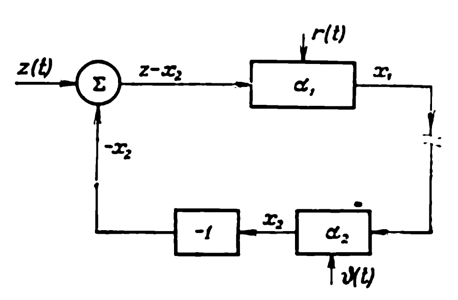


Fig. 2.4

Conformément à (1), les processus qui se produisent dans le système ouvert à la sortie de l'organe α_1 seront décrits par les équations

$$\left.\begin{aligned} f_1(D)\, x_1 &= -e_1(D)\, x_2, \\ f_2(D)\, x_2 &= e_2(D)\, \vartheta(t). \end{aligned}\right\} \qquad (2.25)$$

Il en résulte, compte tenu de (7),

$$\left.\begin{aligned} x_1 &= -W_1(D)\, x_2, \\ x_2 &= W_2(D)\, \vartheta(t) \end{aligned}\right\} \qquad (2.26)$$

ou

$$x_1 = -W_1(D)\,W_2(D)\,\vartheta(t). \tag{2.27}$$

Comme d'après (8)

$$W_1(D)\,W_2(D) = W(D), \tag{2.28}$$

l'équation (27) devient

$$x_1 = -W(D)\,\vartheta(t). \tag{2.29}$$

De la sorte, la matrice $W(D)$ est la *matrice de transfert* prise avec le signe opposé du système commandé *ouvert à la sortie du premier organe.*

Considérons maintenant la possibilité d'interpréter l'opérateur $E_{m_1} + W(D)$. Le signal à la sortie de l'organe α_1, c'est-à-dire le vecteur x_1 après le passage du signal $\vartheta(t)$ par le système ouvert représenté sur la figure 2.4, est défini par l'expression (29). La différence entre le signal $\vartheta(t)$ appliqué à l'entrée de l'organe α_2 et le signal reçu à la sortie du système ouvert sera

$$\vartheta(t) - [-W(D)\,\vartheta(t)] = [E_{m_1} + W(D)]\,\vartheta(t). \tag{2.30}$$

La grandeur (30) a été appelée par Bode différence de retour. Dans ce sens la matrice carrée $E_{m_1} + W(D)$ peut s'appeler *matrice de différence de retour.*

6. Sur la reproduction du signal d'entrée dans un système multivariable. L'équation (17) qui décrit les processus se déroulant dans un système multivariable s'écrit

$$x_1 = [E_{m_1} + W(D)]^{-1}\,W_1(D)\,z(t) + [E_{m_1} + W(D)]^{-1}\,R(D)\,r(t).$$

Comme d'après (8), avec $m_1 = m_2$,

$$W_1(D) = W(D)\,W_2^{-1}(D), \tag{2.31}$$

(17) peut se mettre sous la forme

$$x_1 = [E_{m_1} + W(D)]^{-1}\,W(D)\,W_2^{-1}(D)\,z(t) + \\ + [E_{m_1} + W(D)]^{-1}\,R(D)\,r(t). \tag{2.32}$$

La deuxième équation du système (1) conduit à

$$x_2 = W_2(D)\,x_1. \tag{2.33}$$

En portant dans (33) la valeur de x_1 tirée de (32), on obtient

$$x_2 = W_2(D)\,[E_{m_1} + W(D)]^{-1}\,W(D)\,W_2^{-1}(D)\,z(t) + \\ + W_2(D)\,[E_{m_1} + W(D)]^{-1}\,R(D)\,r(t). \tag{2.34}$$

Pour $r(t) \equiv 0$, l'expression (34) devient

$$x_2 = W_2(D)\,[E_{m_1} + W(D)]^{-1}\,W(D)\,W_2^{-1}(D)\,z(t). \tag{2.35}$$

L'équation (35) est une généralisation de l'équation (1.39) qui décrit la reproduction du signal d'entrée dans un système monovariable.

§ 3. Méthodes fréquentielles d'étude de la stabilité des systèmes commandés linéaires

1. Transformation du déterminant caractéristique d'un système commandé en boucle fermée. Comme il a été montré au § 2 (2.10), le déterminant caractéristique du système en boucle fermée est de la forme

$$\Delta(D) = \Delta_1(D)\,\Delta_2(D) \det [E_{m_1} + W(D)].$$

Ici $\Delta_i(D)$ $(i = 1, 2)$ sont les déterminants caractéristiques des organes du système de commande; d'après (2.21), le produit de ces déterminants est le déterminant du système *en boucle ouverte*

$$\Delta_0(D) = \Delta_1(D)\,\Delta_2(D).$$

Donc le déterminant du système en boucle fermée sera

$$\Delta(D) = \Delta_0(D) \det [E_{m_1} + W(D)]. \tag{3.1}$$

Suivant (2.7) et (2.8)

$$W(D) = W_1(D)\,W_2(D) = \frac{F_1(D)\,e_1(D)\,F_2(D)\,e_2(D)}{\Delta_1(D)\,\Delta_2(D)}, \tag{3.2}$$

par suite

$$W(D) = \frac{F_1(D)\,e_1(D)\,F_2(D)\,e_2(D)}{\Delta_0(D)}. \tag{3.3}$$

La matrice $W(D)$ est une matrice carrée $m_1 \times m_1$ et en vertu de l'expression (3), tous les éléments de la matrice

$$W(D) = \begin{bmatrix} W_{11}(D) & W_{12}(D) & \ldots & W_{1m_1}(D) \\ \cdot & \cdot & \cdot & \cdot \\ W_{m_1 1}(D) & W_{m_1 2}(D) & \ldots & W_{m_1 m_1}(D) \end{bmatrix} \tag{3.4}$$

sont des fonctions rationnelles de l'opérateur de différentiation D, le dénominateur de chacune d'elles étant égal à $\Delta_0(D)$.

La matrice $E_{m_1} + W(D)$ s'écrit

$$E_{m_1} + W(D) = \begin{bmatrix} 1 + W_{11}(D) & W_{12}(D) & \ldots & W_{1m_1}(D) \\ \cdot & \cdot & \cdot & \cdot \\ W_{m_1 1}(D) & W_{m_1 2}(D) & \ldots & 1 + W_{m_1 m_1}(D) \end{bmatrix}. \tag{3.5}$$

Son déterminant peut être ramené à la forme

$$\det [E_{m_1} + W(D)] = 1 + K(D), \tag{3.6}$$

où $K(D)$ est une fonction rationnelle de D:

$$K(D) = \frac{L(D)}{[\Delta_0(D)]^{m_1}}, \tag{3.7}$$

et $L(D)$ un polynôme en D. Il résulte des expressions (6) et (1) que

$$\Delta(D) = \Delta_0(D)\,[1 + K(D)] \tag{3.8}$$

ou d'après (7)

$$\Delta(D) = \Delta_0(D) + \frac{L(D)}{[\Delta_0(D)]^{m_1-1}}. \tag{3.9}$$

$\Delta(D)$ étant une fonction entière (puisque c'est le déterminant du système en boucle fermée), il s'ensuit que

$$L(D) = M(D)\,[\Delta_0(D)]^{m_1-1}, \tag{3.10}$$

où $M(D)$ est un polynôme en D.

Ainsi

$$K(D) = \frac{M(D)}{\Delta_0(D)}. \tag{3.11}$$

L'expression (8) se met maintenant sous la forme

$$\Delta(D) = \Delta_0(D)\left[1 + \frac{M(D)}{\Delta_0(D)}\right] = \Delta_0(D) + M(D). \tag{3.12}$$

Pour qu'un système de commande soit *asymptotiquement stable*, il faut que tous les zéros du polynôme $\Delta(p)$ se situent dans le demi-plan gauche de la variable complexe p.

Les conditions nécessaires et suffisantes pour que ces contraintes soient vérifiées sont données par le critère de Hurwitz [17, 21]. Toutefois, dans ce sens le critère de Nyquist présente des possibilités plus grandes du fait qu'il permet de se faire une idée de la stabilité d'un système *en boucle fermée* d'après les lieux de transfert du système *en boucle ouverte*.

On appelle une caractéristique fréquentielle d'un système en boucle ouverte la fonction

$$W(i\omega) = [W(p)]_{p=i\omega} \quad (-\infty < \omega < \infty),$$

où $W(D)$ est la matrice de transfert du système ouvert, définie au § 2 par la formule (2.28).

Puisque d'après (2.28)

$$W(D) = W_1(D)\,W_2(D),$$

où $W_1(D)$ et $W_2(D)$ sont les matrices de transfert des éléments du système, on a

$$W(i\omega) = W_1(i\omega)\,W_2(i\omega),$$

et pour construire le lieu de transfert d'un système en boucle ouverte il ne faut connaître que les lieux de transfert de ses organes. Or, si l'on ne connaît pas l'expression analytique des fonctions $W_i(D)$, les lieux de transfert en question peuvent être établis par expérience de la façon décrite dans ce qui suit. Ainsi l'importance

principale du critère de Nyquist consiste dans le fait qu'il permet de juger sur la stabilité en boucle fermée non seulement d'après les caractéristiques analytiques mais aussi d'après les caractéristiques empiriques du système *en boucle ouverte*. Ce critère, dont nous donnons la déscription ci-dessous, a été proposé par Nyquist [13] pour les amplificateurs à réaction et appliqué par A. Mikhaïlov [65] et autres aux systèmes réglables.

2. Critère de stabilité asymptotique des systèmes en boucle fermée ne comportant que des éléments stables (critère de Nyquist). Examinons d'abord les systèmes commandés dont les éléments sont stables. Cela signifie que les déterminants caractéristiques des éléments $\Delta_i(p)$ n'ont pas de zéros dans le demi-plan droit de la variable complexe p.

Nous allons nous borner à l'examen des systèmes tels que le degré du numérateur de la fonction

$$K(D) = \frac{M(D)}{\Delta_0(D)}$$

ne soit pas supérieur au degré du dénominateur. Comme d'après (8)

$$\Delta(p) = \Delta_0(p)[1 + K(p)] = \Delta_0(p)\left[1 + \frac{M(p)}{\Delta_0(p)}\right],$$

on voit bien que la fonction $\Delta(p)$ a pour zéros les zéros de la fonction

$$1 + K(p) = \frac{\Delta_0(p) + M(p)}{\Delta_0(p)}. \tag{3.13}$$

Les zéros de la fonction $1 + K(p)$ sont formés par les zéros de son numérateur, et les pôles de cette fonction, par les zéros de son dénominateur. Si tous les organes du système sont asymptotiquement stables, la fonction $\Delta_0(p)$ ne possède pas de zéros dans le demi-plan droit de la variable complexe p. Pour cette raison la valeur de l'intégrale qui suit, prise suivant le contour fermé C situé dans le demi-plan droit de p, est donnée par l'expression

$$\frac{1}{2\pi i}\oint_C \frac{K'(p)\,dp}{1 + K(p)} = -N + P = -N, \tag{3.14}$$

où N est le nombre de zéros et P le nombre de pôles que la fonction $1 + K(p)$ possède dans le domaine délimité par le contour C.

Remarquons que généralement, en énonçant le théorème de Cauchy, on recourt dans la théorie des fonctions de la variable complexe à la rotation suivant le contour extérieur fermé s'effectuant dans le sens inverse à celui des aiguilles d'une montre, c'est-à-dire de façon que le domaine délimité par ce contour reste à gauche. Comme nous avons adopté ici le parcours dans le sens des aiguilles d'une montre, le second membre de l'expression (14) s'écrit $-N + P$.

Si à l'aide de la représentation conforme

$$K = K(p)$$

on applique le contour C sur le plan K, l'expression (14) devient

$$\frac{1}{2\pi i}\oint_{\Gamma}\frac{dK}{1+K} = -N, \tag{3.15}$$

où Γ est le contour dans le plan K obtenu par représentation conforme. Le sens de parcours le long de ce contour est défini par l'application et ne peut être choisi arbitrairement.

Puisque les coefficients des polynômes $M(p)$ et $\Delta_0(p)$ sont réels, aux points réels p correspondent les points réels du plan K. Au point $p = 0$ et au point à l'infini $p = \infty$ correspondent également les points réels $K(0)$ et $K(\infty)$. Aux points $p = \varepsilon + i\omega$ et $p = \varepsilon - i\omega$ correspondent les points symétriques $K(\varepsilon + i\omega)$ et $K(\varepsilon - i\omega) = \overline{K}(\varepsilon + i\omega)$. Donc, si le contour C est symétrique par rapport à l'axe réel, le contour Γ situé dans le plan K le sera également. La grandeur $1 + K$ est une grandeur complexe. Dans le plan K elle représente le rayon vecteur d'un point situé sur le contour Γ. L'origine de ce rayon vecteur se trouve au point $(-1, 0)$.

Puisque

$$1+K=|1+K|\,e^{i\arg(1+K)}, \tag{3.16}$$

$$\ln(1+K)=\ln|1+K|+i\arg(1+K), \tag{3.17}$$

$$\frac{dK}{1+K}=d\ln(1+K)=d\ln|1+K|+i\,d\arg(1+K), \tag{3.18}$$

il vient

$$-N=\frac{1}{2\pi i}\oint_{\Gamma}\frac{dK}{1+K}=\frac{1}{2\pi}\oint_{\Gamma}d\arg(1+K)=\frac{\Delta\arg(1+K)}{2\pi}, \tag{3.19}$$

où $\Delta\arg(1 + K)$ est l'accroissement de l'argument du rayon vecteur $1 + K$ lorsqu'on parcourt le contour fermé Γ. L'intégrale du premier terme de l'expression (18)

$$\oint_{\Gamma}d\ln|1+K|=0, \tag{3.20}$$

puisque le contour Γ est fermé.

Ainsi, N est égal au nombre S de tours complets du vecteur $1 + K$ autour du point $(-1, 0)$, lorsque p décrit le contour C dans le sens des aiguilles d'une montre, les rotations du vecteur $1 + K$ étant également considérées positives dans le sens des aiguilles d'une montre

$$S = N. \tag{3.21}$$

Remarquons que si l'on considère les rotations du vecteur $1 + K$ positives lorsqu'elles s'effectuent dans le sens opposé à celui des aiguilles d'une montre, alors, d'après (15), le nombre de tours du vecteur $1 + K$ sera égal à $-N$.

Pour résoudre le problème relatif à la stabilité asymptotique d'un système en boucle fermée, il faut choisir dans le plan p un contour encerclant le demi-plan droit tout entier.

Supposons que le contour dans le plan p soit composé d'un demi-cercle L de rayon R centré à l'origine des coordonnées et de son diamètre $C' + C$ reposant sur un axe imaginaire (fig. 3.1). Le sens de parcours est celui des aiguilles d'une montre.

Quand p varie le long de l'axe imaginaire de 0 à $+i\infty$, le point $K(p)$ parcourt le trajet Γ du point $K(0)$ au point $K(\infty)$ situé sur

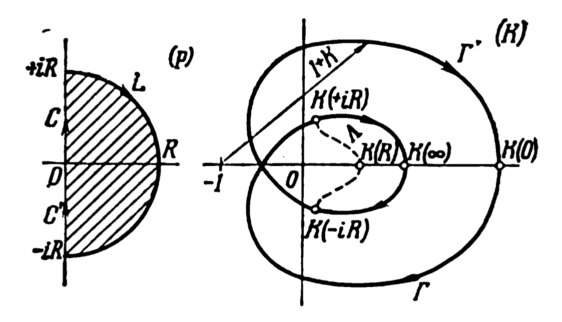


Fig. 3.1

l'axe réel. La symétrique Γ' de ce trajet par rapport à l'axe réel permet d'obtenir le contour fermé $\Gamma + \Gamma'$ matérialisé par le trait plein.

Au demi-cercle L de $p = +iR$ à $p = -iR$ en passant par $p = R$ correspond l'arc Λ de $K = K(+iR)$ à $K = K(-iR)$ passant par $K = K(R)$ sur l'axe réel. Cet arc Λ est une bifurcation du contour $\Gamma' + \Gamma$.

Avec $R \to \infty$ le contour dans le plan p entoure à la limite le demi-plan droit tout entier. On voit aisément que si $R \to \infty$ la ramification Λ se trouve contractée au point $K(\infty)$. En effet, pour $p = Re^{i\varphi}$, nous aurons

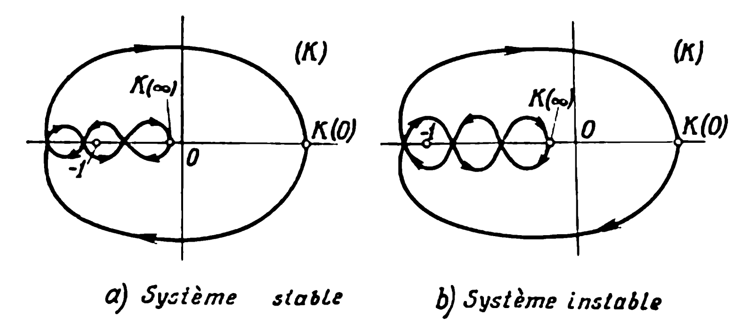
$$K(p) = \frac{M(p)}{\Delta_0(p)} = \frac{a_0 R^n e^{i\varphi n} + a_1 R^{n-1} e^{i\varphi(n-1)} + \ldots + a_n}{b_0 R^n e^{i\varphi n} + b_1 R^{n-1} e^{i\varphi(n-1)} + \ldots + b_n}, \quad (3.22)$$

$$\lim_{R \to \infty} K(p) = K(\infty) = \frac{a_0}{b_0}. \quad (3.23)$$

Si le degré du polynôme $M(p)$ est inférieur à celui de $\Delta_0(p)$, c'est-à-dire si $a_0 = 0$, on a $K(\infty) = 0$. Ainsi, avec $R \to \infty$, il ne reste sur le plan K que le contour fermé $\Gamma + \Gamma'$.

Le nombre S de tours du contour $\Gamma + \Gamma'$ autour du point (—1, 0) dans le sens des aiguilles d'une montre est égal au nombre de zéros N du déterminant caractéristique Δ (p) dans le demi-plan droit p. Si le nombre de tours est nul, le système en boucle fermée est asymptotiquement stable.

En particulier, si le contour $\Gamma + \Gamma'$ n'a pas de points doubles, la stabilité asymptotique n'a lieu que dans le cas où le point (—1, 0)



a) Système stable *b) Système instable*

Fig. 3.2

repose au-dehors du contour. C'est ce qui détermine le *critère de Nyquist* [17].

Constatons que des cas sont possibles où le contour $\Gamma + \Gamma'$ embrasse le point (—1, 0) mais les points doubles (fig. 3.2) font que le nombre de rotations du contour autour du point (—1, 0) est nul. Dans ces cas le système est stable.

Signalons encore la position relative des points K (0) et K (∞) dans le plan K. Il est facile de voir que les nombres Δ_0 (0) et Δ_0 (∞) ont toujours les mêmes signes, sinon le polynôme Δ_0 (p) aurait au moins une racine réelle positive, ce qui contredit l'hypothèse sur la stabilité d'un système en boucle ouverte. Si un système en boucle fermée est stable, les nombres $\Delta_0 (0) + M (0)$, $\Delta_0 (\infty) + M (\infty)$ doivent être de même signe, d'où il s'ensuit que les signes des nombres réels $1 + K (0)$, $1 + K (\infty)$ doivent être les mêmes. Ainsi, dans un système *stable*, les points K (0) et K (∞) reposent du même côté du point (—1, 0). Dans un système *instable*, les points K (0) et K (∞) peuvent se trouver soit d'un côté, soit des deux côtés différents par rapport au point (—1, 0).

3. Application du critère de Nyquist aux systèmes à organes neutres. Les objets à asservir peuvent être neutres ou instables (machine à vapeur, navire, avion neutre ou statiquement instable) et le problème d'asservissement consiste à les rendre stables. Parmi les éléments d'un régulateur il existe également des organes neutres (servo-moteurs). En présence des éléments neutres le point $p = 0$ est le pôle de la fonction K (p), et en construisant

un contour dans le plan p, ce pôle doit être contourné suivant le petit demi-cercle l de rayon r (fig. 3.3).

Si le point $p = 0$ est un pôle d'ordre un, dans son voisinage la fonction $K(p)$ se comporte comme $\frac{1}{p}$, et aux arguments $+i0$, $-i0$ de la fonction $K(i\omega)$ correspondent les branches des courbes Γ, Γ', qui s'éloignent à l'infini le long de l'axe imaginaire. Le demi-cercle l

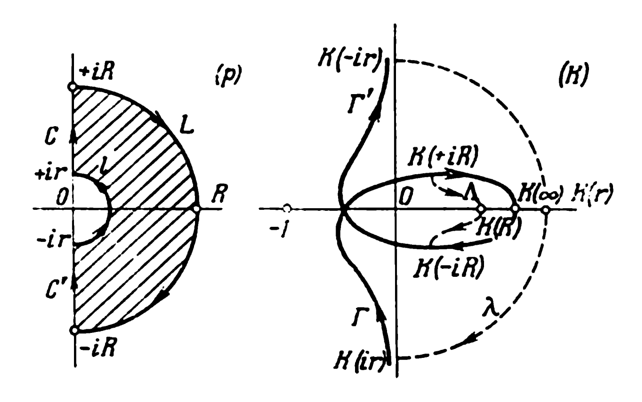


Fig. 3.3

a pour image une courbe voisine d'un demi-cercle de rayon infini décrite dans le sens des aiguilles d'une montre et qui se joint aux branches infinies des courbes Γ, Γ'. Lorsque $R \to \infty$, $r \to 0$, l'arc Λ qui est l'image du demi-cercle L se contracte en un point $K(\infty)$, et l'arc λ qui est l'image du petit demi-cercle l s'éloigne à l'infini.

Finalement, il ne reste de nouveau que la courbe $\Gamma + \Gamma'$ dont les branches infinies doivent être supposées fermées par un demi-cercle de rayon infiniment grand, situé dans le demi-plan K droit. Au contour fermé ainsi obtenu on peut également appliquer le critère de Nyquist.

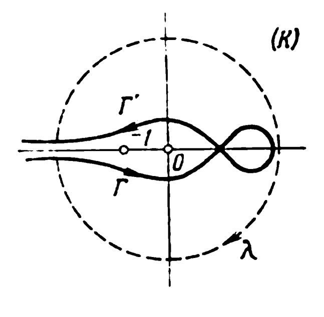


Fig. 3.4

Le lieu de Nyquist de la figure 3.3 se rapporte à un système stable dont le point $p = 0$ est un pôle de premier ordre. Le cas de la figure 3.3 vérifie la relation

$$S + \frac{1}{2} = N. \tag{3.24}$$

Le lieu de Nyquist représenté sur la figure 3.4 se rapporte à un

système stable pour lequel le point $p = 0$ est un pôle de deuxième ordre.

Si le point $p = 0$ est un pôle d'ordre n, la courbe $\Gamma + \Gamma'$ doit être fermée par une ligne éloignée à l'infini comportant n demi-cercles décrits dans le sens des aiguilles d'une montre. Cela résulte du fait que dans le voisinage de $p = 0$ la fonction $K(p)$ se comporte comme $\frac{1}{p^n} = \frac{1}{r^n e^{i\varphi n}}$, c'est-à-dire que dans le plan K l'angle polaire est n fois plus grand que son homologue dans le plan p. Dans le cas envisagé on a

$$S + \frac{n}{2} = N. \tag{3.25}$$

4. Application du critère de Nyquist aux systèmes à organes instables. En présence d'un organe instable, la fonction $K(p)$ possède un pôle d'ordre de multiplicité n au point $p = a$ où $\operatorname{Re} a > 0$. Supposons pour fixer les idées que le point a repose sur l'axe réel. En construisant le contour dans le plan p, faisons exclure le pôle a à l'aide du cercle l de petit rayon r, décrit dans le sens opposé à celui des aiguilles d'une montre. La courbe λ qui lui correspond dans le plan K est voisine d'un cercle de rayon infiniment grand, décrit n fois dans le sens des aiguilles d'une montre.

Pour cette raison, en appliquant le critère de Nyquist, il faut ajouter au contour $\Gamma + \Gamma'$ un cercle infiniment grand décrit n fois dans le sens des aiguilles d'une montre. Ceci résulte du fait que dans le

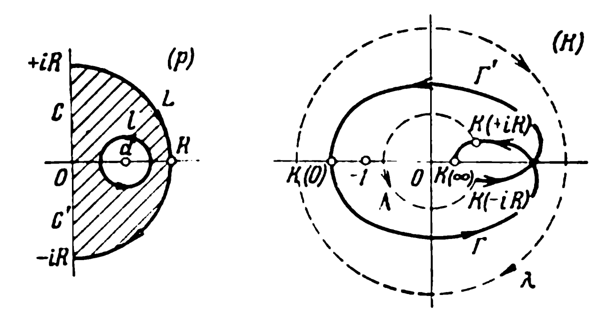


Fig. 3.5

voisinage du pôle $p = a$ la fonction $K(p)$ se comporte comme

$$\frac{1}{(p-a)^n} = \frac{1}{r^n e^{i\varphi n}}. \tag{3.26}$$

La présence d'un organe instable conduit à la relation

$$S + n = N. \tag{3.27}$$

Le lieu de Nyquist représenté sur la figure 3.5 se rapporte à un système stable dont le point $p = a$ est un pôle de premier ordre $(n = 1)$.

Dans tous les cas examinés aux points 3 et 4, la formule

$$N = S + P \tag{3.28}$$

s'avère valable si l'on prend les pôles simples situés aux points intérieurs du demi-plan droit pour une unité et ceux qui se situent sur l'axe imaginaire pour $\frac{1}{2}$, les pôles multiples respectivement pour n et $\frac{n}{2}$.

5. Lieux de transfert des systèmes commandés et leur détermination expérimentale. Comme nous l'avons montré au point 2, l'application du critère de Nyquist dans le cas de la construction d'une représentation conforme $K = K(p)$ n'exige réellement que la construction de la

$\vartheta(t)$ → [$\Psi(D)$] → x

Fig. 3.6

représentation $K(i\omega)$ $(-\infty < \omega < \infty)$. Or, comme d'après (6)

$$1 + K(i\omega) = \det[E_{m_1} + W(i\omega)] = \det[E_{m_1} + W_1(i\omega) W_2(i\omega)], \tag{3.29}$$

pour pouvoir appliquer le critère de Nyquist, il ne faut connaître que les lieux de transfert de $W_s(i\omega)$ $(s = 1, 2)$ des organes du système. Puisque $W_s(D)$ $(s = 1, 2)$ est une fonction rationnelle de D dont les coefficients sont réels, $W_s(-i\omega) = \overline{W_s(i\omega)}$; il suffit donc de définir la fonction $W_s(i\omega)$ seulement sur l'intervalle $0 < \omega < \infty$.

Voyons maintenant s'il est possible de déterminer expérimentalement le lieu de transfert d'un organe décrit par une équation différentielle vectorielle du type

$$f(D)\, x = e(D)\, \vartheta(t). \tag{3.30}$$

Soient $f(D)$ la matrice $n \times n$, x le vecteur de dimension n, $e(D)$ la matrice $n \times m$, $\vartheta(t)$ le vecteur de dimension m.

L'équation (30) entraîne

$$x = \Psi(D)\, \vartheta(t), \tag{3.31}$$

où

$$\Psi(D) = \frac{F(D)\, e(D)}{\Delta(D)}. \tag{3.32}$$

La matrice de transfert $\Psi(D)$ de l'organe considéré (fig. 3.6) est une matrice $n \times m$.

Soit le signal d'entrée $\vartheta(t)$

$$\vartheta(t) = A\cos\omega_1 t = \frac{1}{2}(Ae^{i\omega_1 t} + Ae^{-i\omega_1 t}), \tag{3.33}$$

où A est un vecteur de dimension m. L'équation différentielle vectorielle (30) pourra maintenant s'écrire

$$f(D)\,x = \frac{1}{2}e(D)(Ae^{i\omega_1 t} + Ae^{-i\omega_1 t}). \tag{3.34}$$

En admettant que $i\omega_1$ n'est pas une racine de l'équation caractéristique $\det f(p) = 0$, cherchons une solution particulière de l'équation (34) sous la forme

$$\tilde{x} = \frac{1}{2}(h_1 e^{i\omega_1 t} + h_2 e^{-i\omega_1 t}). \tag{3.35}$$

Portons l'expression (35) dans (34) pour trouver

$$h_1 = \Psi(i\omega_1)A, \quad h_2 = \Psi(-i\omega_1)A. \tag{3.36}$$

Ainsi,

$$\tilde{x} = \frac{1}{2}[\Psi(i\omega_1)Ae^{i\omega_1 t} + \Psi(-i\omega_1)Ae^{-i\omega_1 t}]. \tag{3.37}$$

L'élément $\tilde{x}_j$ du vecteur $\tilde{x}$ peut se mettre sous la forme:

$$\tilde{x}_j = \frac{1}{2}\left[\sum_{k=1}^{m}\Psi_{jk}(i\omega_1)A_k e^{i\omega_1 t} + \sum_{k=1}^{m}\Psi_{jk}(-i\omega_1)A_k e^{-i\omega_1 t}\right]. \tag{3.38}$$

Désignons par $R_{jk}(\omega_1)$ et $\psi_{jk}(\omega_1)$ le module et l'argument de la grandeur complexe $\Psi_{jk}(i\omega_1)$:

$$\Psi_{jk}(i\omega_1) = R_{jk}(\omega_1)e^{i\psi_{jk}(\omega_1)}. \tag{3.39}$$

Comme nous l'avons indiqué plus haut, la fonction $\Psi_{jk}(i\omega)$ est le lieu de transfert ou le *lieu de réponse en fréquence*; la fonction $R_{jk}(\omega)$ est une réponse fréquentielle *en amplitude*; la fonction $\psi_{jk}(\omega)$, une réponse fréquentielle *en phase*. La fonction

$$P_{jk}(\omega) = \operatorname{Re}\Psi_{jk}(i\omega)$$

se nomme réponse en fréquence *réelle* et la fonction

$$Q_{jk}(\omega) = \operatorname{Im}\Psi_{jk}(i\omega),$$

réponse en fréquence *imaginaire*.

En vertu de (39), la solution particulière de (38) peut s'écrire

$$\tilde{x}_j = \sum_{k=1}^{m} R_{jk}(\omega_1)A_k\cos[\omega_1 t + \psi_{jk}(\omega_1)] \qquad (j = 1, \ldots, n). \tag{3.40}$$

D'après l'exposé ci-dessus, la matrice

$$\Psi(i\omega) = [\Psi_{jk}(i\omega)] = [R_{jk}(\omega)\, e^{i\psi_{jk}(\omega)}] \tag{3.41}$$

est la *matrice* de réponse en fréquence de l'organe étudié.

Décrivons maintenant une expérience qui permet de déterminer la matrice $\Psi(i\omega)$.

Soit le vecteur A figurant dans l'expression (33) dont un seul élément $A_1 \neq 0$, alors que $A_2 = A_3 = \ldots = A_m = 0$. L'expression (40) devient alors

$$\tilde{x}_j = R_{j1}(\omega_1) A_1 \cos[\omega_1 t + \psi_{j1}(\omega_1)] \qquad (j = 1, \ldots, n).$$

Après avoir mesuré le signal à la sortie du système (après l'amortissement des oscillations propres du système), trouvons les valeurs de $R_{j1}(\omega_1)$ et $\psi_{j1}(\omega_1)$ $(j = 1, \ldots, n)$. En reprenant l'expérience avec $\vartheta(t) = A \cos \omega_2 t$, $\vartheta(t) = A \cos \omega_3 t$, etc., on peut construire les courbes des fonctions $R_{j1}(\omega)$ et $\psi_{j1}(\omega)$ $(j = 1, \ldots, n)$. Ainsi, à l'aide de l'expérience décrite nous trouverons les éléments de la première colonne de la matrice (41), c'est-à-dire les fonctions

$$\Psi_{11}(i\omega),$$
$$\Psi_{21}(i\omega),$$
$$\cdot\;\cdot\;\cdot\;\cdot\;\cdot$$
$$\Psi_{n1}(i\omega).$$

En procédant maintenant à la deuxième série d'expériences qui ne diffère de la première que par $A_2 \neq 0$, alors que $A_1 = A_3 = A_4 = \ldots = A_m = 0$, nous pourrons, d'après les données de mesure, déterminer les éléments de la deuxième colonne de la matrice (41), c'est-à-dire les fonctions

$$\Psi_{12}(i\omega),$$
$$\Psi_{22}(i\omega),$$
$$\cdot\;\cdot\;\cdot\;\cdot\;\cdot$$
$$\Psi_{n2}(i\omega).$$

Après m séries d'expériences de ce type, on obtient toute la matrice $\Psi(i\omega)$

$$\Psi(i\omega) = \begin{bmatrix} \Psi_{11}(i\omega) & \Psi_{12}(i\omega) & \ldots & \Psi_{1m}(i\omega) \\ \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot \\ \Psi_{n1}(i\omega) & \Psi_{n2}(i\omega) & \ldots & \Psi_{nm}(i\omega) \end{bmatrix}.$$

Les lieux de transfert des organes isolés (et, par suite, du système en boucle ouverte tout entier), nécessaires pour appliquer le critère de Nyquist, peuvent être obtenus par expérience. Ceci est très im-

portant lorsque la description analytique complète de certains éléments ou les valeurs de certains paramètres des éléments, etc., sont inconnues. Toutefois, on doit être sûr que l'élément considéré est décrit par les équations différentielles linéaires à coefficients constants.

6. Exemple de construction du lieu de Nyquist. Examinons le système décrit par les équations différentielles

$$\left.\begin{aligned} Dx_1 + x_1 &= -\mu_2 [z(t) - x_2], \\ Dx_2 + x_2 &= \mu_1 x_1. \end{aligned}\right\} \tag{3.42}$$

Selon les notations du § 2, les équations (2.1) deviennent

$$\left.\begin{aligned} &f_1(D) = D+1, \qquad e_1(D) = -\mu_2, \\ &f_2(D) = D+1, \qquad e_2(D) = \mu_1, \\ &W_1(D) = -\frac{\mu_2}{D+1}, \qquad W_2(D) = \frac{\mu_1}{D+1}, \\ &W(D) = W_1(D)\, W_2(D) = -\frac{\mu_1\mu_2}{(D+1)^2}. \end{aligned}\right\} \tag{3.43}$$

Dans cet exemple $m_1 = 1$, c'est-à-dire la fonction $W(D)$ est un scalaire, et donc

$$\det [E_{m_1} + W(D)] = 1 + W(D). \tag{3.44}$$

Par conséquent, la fonction $K(D)$, déterminée d'après (6), se met

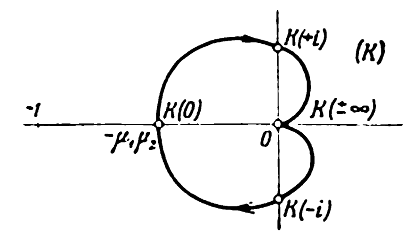


Fig. 3.7

ici sous la forme

$$K(D) = W(D) = -\frac{\mu_1\mu_2}{(D+1)^2}. \tag{3.45}$$

Il en résulte que

$$K(i\omega) = -\frac{\mu_1\mu_2}{(i\omega+1)^2} = -\mu_1\mu_2 \frac{1-\omega^2-2i\omega}{(1+\omega^2)^2}, \tag{3.46}$$

$$|K(i\omega)| = \frac{\mu_1\mu_2}{1+\omega^2}, \quad K(\infty) = 0, \quad K(0) = -\mu_1\mu_2. \tag{3.47}$$

Le lieu de Nyquist de l'exemple examiné est donné par la figure 3.7. Cette figure montre que le système (42) sera stable si on respecte la

condition

$$\mu_1\mu_2 < 1. \tag{3.48}$$

La condition de stabilité (48) coïncide évidemment avec celle obtenue à partir de l'équation caractéristique du système des équations différentielles (42):

$$\Delta(p) = \begin{vmatrix} p+1 & -\mu_2 \\ -\mu_1 & p+1 \end{vmatrix} = p^2+2p+1-\mu_1\mu_2 = 0. \tag{3.49}$$

7. Systèmes à organes de retard et leurs critères de stabilité. Examinons un système monovariable (fig. 3.8) décrit par les équations différentielles scalaires

$$\left.\begin{aligned} f(D)\,x(t) &= g(D)\,[z(t) - x_1(t)], \\ x_1(t) &= x(t-\tau) \quad (\tau = \text{const}), \end{aligned}\right\} \tag{3.50}$$

où $f(D)$ et $g(D)$ sont des polynômes (à coefficients constants) par rapport à l'opérateur de dérivation $D = d/dt$:

$$\left.\begin{aligned} f(D) &= a_0D^n + a_1D^{n-1} + \ldots + a_{n-1}D + a_n, \\ g(D) &= b_0D^m + b_1D^{m-1} + \ldots + b_{m-1}D + b_m \quad (m < n). \end{aligned}\right\} \tag{3.51}$$

La première équation (50) décrit les processus ayant lieu dans le premier organe du système (fig. 3.8). La fonction de transfert $G(D)$ de cet organe est

$$G(D) = \frac{g(D)}{f(D)}. \tag{3.52}$$

La deuxième équation(50) décrit les processus qui se déroulent dans le deuxième organe dans lequel a lieu le retard du signal d'une valeur

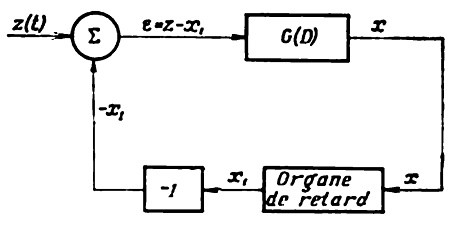


Fig. 3.8

τ = const. Cet organe peut se nommer organe de retard. Le système d'équations (50) peut être remplacé par une équation différentielle-fonctionnelle équivalente

$$f(D)\,x(t) + g(D)\,x(t-\tau) = g(D)\,z(t). \tag{3.53}$$

Les équations différentielles-fonctionnelles constituent une classe très importante des équations fonctionnelles. La théorie des équations différentielles-fonctionnelles a été élaborée par L. Pontriaguine,

A. Mychkis, R. Bellman, N. Krassovski, L. Elsgoltz, V. Zoubov et d'autres savants [11, 31, 43, 67, 92].

Les oscillations libres d'un système en boucle fermée (fig. 3.8) seront décrites par une équation différentielle-fonctionnelle homogène

$$f(D)\,x(t) + g(D)\,x(t-\tau) = 0, \tag{3.54}$$

tirée de l'équation (53) en y posant $z(t) \equiv 0$.

La solution particulière de l'équation (54) sera recherchée sous la forme

$$x(t) = C_i e^{\gamma_i t}, \tag{3.55}$$

où $\gamma_i = \text{const}$. Puisque d'après (55)

$$x(t-\tau) = C_i e^{-\gamma_i \tau} e^{\gamma_i t},$$

$$\frac{d^s}{dt^s}[x(t-\tau)] = C_i e^{-\gamma_i \tau} \gamma_i^s e^{\gamma_i t} \quad (s = 1, 2, \ldots),$$

après le report de (55) dans (54), on obtient

$$f(\gamma_i)\,C_i e^{\gamma_i t} + e^{-\gamma_i \tau} g(\gamma_i)\,C_i e^{\gamma_i t} = 0. \tag{3.56}$$

Etant donné que $C_i e^{\gamma_i t} \neq 0$, il résulte de l'équation (56) que γ_i doit être la racine de l'équation

$$\Delta(p) = 0, \tag{3.57}$$

où

$$\Delta(p) = f(p) + e^{-p\tau} g(p). \tag{3.58}$$

(57) est l'équation caractéristique de l'équation différentielle-fonctionnelle (54). D'après (58), l'équation caractéristique (57) est une équation transcendante.

Aux racines multiples γ_s de l'équation caractéristique (57) correspondent non seulement la solution $e^{\gamma_s t}$ mais encore les solutions

$$te^{\gamma_s t}, \; t^2 e^{\gamma_s t}, \; \ldots, \; t^{q_s - 1} e^{\gamma_s t},$$

où q_s est l'ordre de multiplicité de la racine γ_s.

Pour les équations différentielles-fonctionnelles linéaires de la forme (54) à coefficients constants et à retard constant (τ = const), telles que toutes les racines de l'équation caractéristique (57) possèdent les parties réelles négatives, on a démontré [31] que toute solution peut être développée en une série absolument et uniformément convergente composée de solutions fondamentales

$$\left.\begin{aligned} x(t) &= \sum_{s=1}^{\infty} P_s(t)\, e^{\lambda_s t}, \\ Re\lambda_1 \geqslant Re\lambda_2 &\geqslant \ldots \geqslant Re\lambda_n \geqslant \ldots, \end{aligned}\right\} \tag{3.59}$$

où $P_s(t)$ sont des polynômes en t de degré égal ou supérieur à $q_s - 1$. Pour ces mêmes équations on a démontré également [31] qu'avec $t > t_0 + \tau$, on a

$$\left| \sum_{s=n+1}^{\infty} P_s(t)\, e^{\lambda_s t} \right| \leqslant B e^{(\lambda_{n+1}+\varepsilon)t}, \tag{3.60}$$

où B et ε sont des constantes, et ε est de plus un nombre positif aussi petit que l'on veut.

De ce qui vient d'être dit il résulte que pour faire tendre asymptotiquement vers zéro toutes les solutions $x(t)$ de l'équation (54) avec $t \to \infty$, il faut et il suffit que toutes les racines de l'équation caractéristique (57) aient leurs parties réelles négatives.

Ainsi, la condition de stabilité asymptotique d'un système en boucle fermée (54) muni d'organes de retard consiste dans le fait que le quasi-polynôme (58) qu'on peut écrire sous la forme

$$\Delta(p) = f(p)\,[1 + K(p)], \tag{3.61}$$

avec

$$K(p) = e^{-p\tau} \frac{g(p)}{f(p)} = e^{-p\tau} G(p), \tag{3.62}$$

ne doit pas posséder de zéros dans le demi-plan droit de la variable complexe p.

Comme il résulte de la démonstration du critère de Nyquist, la détermination du nombre de zéros du quasi-polynôme $\Delta(p)$, situés dans le demi-plan droit de la variable complexe p, peut s'effectuer dans le problème considéré d'après ce même critère exposé ci-dessus aux points 2, 3 et 4. Cette possibilité a été montrée par I. Cypkin. Conformément à l'expression (62)

$$K(i\omega) = e^{-i\omega\tau} G(i\omega). \tag{3.63}$$

Puisque d'après (63)

$$K(i\omega) = |K(i\omega)|\, e^{i \arg K(i\omega)} = |G(i\omega)|\, e^{i[\arg G(i\omega) - \omega\tau]},$$

pour tout $\omega = \omega_1$ le vecteur $K(i\omega_1)$ sera égal en module au vecteur $G(i\omega_1)$, mais fera avec lui un angle $\omega_1\tau$ dans le sens des aiguilles d'une montre (fig. 3.9). Dans le domaine de faibles valeurs de ω, les hodographes des vecteurs $K(i\omega)$ et $G(i\omega)$ diffèrent peu mais avec l'accroissement de ω la différence $\omega\tau$ des arguments de $K(i\omega)$ et de $G(i\omega)$ devient de plus en plus importante. Pour $\omega \to \infty$, l'hodographe $K(i\omega)$ s'enroule asymptotiquement sur le point $K(\infty)$ reposant sur l'axe réel.

S'il existe des valeurs de ω telles que $|G(i\omega)| > 1$, alors même dans le cas où l'hodographe du vecteur $G(i\omega)$ n'entoure pas le point $(-1, 0)$, c'est-à-dire celui d'un système stable sans retard, il devient

possible que le point (—1, 0) soit entouré par l'hodographe du vecteur $K(i\omega)$. Autrement dit, des cas sont possibles où l'introduction d'un organe de retard dans un système en boucle fermée stable transforme ce dernier en système instable.

La construction efficace des solutions des équations différentielles-fonctionnelles est une tâche suffisamment ardue. Dans certains

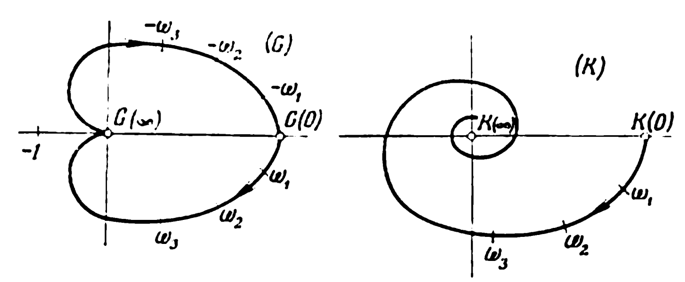


Fig. 3.9

cas il est rationnel d'appliquer les méthodes de calcul opérationnel. A titre de tel exemple examinons l'équation

$$\ddot{x}(t) + a_1\dot{x}(t) + a_2 x(t) = b[z(t) - x(t-\tau)] \quad (\tau = \text{const}). \tag{3.64}$$

Désignons par $\zeta(p)$ et $\xi(p)$ les transformées des fonctions $z(t)$ et $x(t)$:

$$\zeta(p) \doteqdot z(t), \quad \xi(p) \doteqdot x(t).$$

Comme

$$\ddot{x}(t) \doteqdot p^2\xi(p) - p^2x(0) - p\dot{x}(0), \quad \dot{x}(t) \doteqdot p\xi(p) - px(0),$$

$$x(t-\tau) \doteqdot e^{-p\tau}\xi(p) + e^{-p\tau}p\int_{-\tau}^{0} x(t)e^{-pt}\,dt,$$

la transformée de la solution cherchée $x(t)$ est

$$\xi(p) = p\,\frac{x(0)\,p + a_1x(0) + \dot{x}(0)}{f(p) + be^{-p\tau}} - p\,\frac{be^{-p\tau}}{f(p) + be^{-p\tau}}\int_{-\tau}^{0} x(t)e^{-pt}\,dt + \frac{b\zeta(p)}{f(p) + be^{-p\tau}}, \tag{3.65}$$

où

$$f(p) = p^2 + a_1p + a_2.$$

Pour trouver la solution, il faut chercher l'original dont la transformée est $\xi(p)$. On voit de (65) que la solution cherchée dépend de la

variation de la fonction $x(t)$ sur le segment $-\tau \leqslant t \leqslant 0$. Pour rendre le problème résoluble, la fonction

$$\varphi(t) \equiv x(t) \quad (-\tau \leqslant t \leqslant 0), \tag{3.66}$$

dite *fonction initiale*, doit être donnée à l'avance. La solution de (64) qui satisfait à la condition (66) est notée $x_\varphi(t)$.

8. Courbes logarithmiques. La fonction de transfert d'un système en boucle ouverte monovariable décrit par l'équation différentielle linéaire

$$f(D)\,x = g(D)\,z(t) \quad \left(D = \frac{d}{dt}\right), \tag{3.67}$$

où

$$\left.\begin{aligned} f(D) &= a_0D^n + a_1D^{n-1} + \ldots + a_{n-1}D + a_n, \\ g(D) &= b_0D^m + b_1D^{m-1} + \ldots + b_{m-1}D + b_m \quad (m<n), \end{aligned}\right\} \tag{3.68}$$

s'écrit

$$G(D) = \frac{b_0D^m + b_1D^{m-1} + \ldots + b_{m-1}D + b_m}{a_0D^n + a_1D^{n-1} + \ldots + a_{n-1}D + a_n}. \tag{3.69}$$

Pour $z(t) \equiv 1(t)$ l'équation différentielle (67) possède une solution particulière

$$\tilde{x} = K_0, \tag{3.70}$$

avec

$$K_0 = \frac{b_m}{a_n}.$$

La grandeur K_0 s'appelle *gain en boucle ouverte*.

Introduisons les notations

$$G_0(D) = \frac{\dfrac{b_0}{b_m}D^m + \dfrac{b_1}{b_m}D^{m-1} + \ldots + \dfrac{b_{m-1}}{b_m}D + 1}{\dfrac{a_0}{a_n}D^n + \dfrac{a_1}{a_n}D^{n-1} + \ldots + \dfrac{a_{n-1}}{a_n}D + 1}, \tag{3.71}$$

et mettons la fonction de transfert du système en boucle ouverte sous la forme

$$G(D) = K_0G_0(D). \tag{3.72}$$

Nous avons déjà dit que la fonction $G(i\omega) = [G(p)]_{p=i\omega}$ s'appelle réponse en fréquence en boucle ouverte. En désignant par $R(\omega)$ le module de la fonction $G(i\omega)$ et par $\psi(\omega)$, l'argument de cette fonction, on a

$$G(i\omega) = R(\omega)\,e^{i\psi(\omega)} \tag{3.73}$$

ou

$$G(i\omega) = K_0R_0(\omega)\,e^{i\psi(\omega)}, \tag{3.74}$$

avec

$$R_0(\omega) = |G_0(i\omega)|. \qquad (3.75)$$

En calculant les logarithmes des premier et second membres de l'expression (74) on obtient

$$\lg G(i\omega) = \lg K_0 + \lg R_0(\omega) + i\psi(\omega) \lg e. \qquad (3.76)$$

On appelle *réponse en amplitude logarithmique* la fonction

$$L(\omega) = 20 \lg R(\omega) = 20 \lg K_0 + 20 \lg R_0(\omega). \qquad (3.77)$$

Comme argument de la fonction $L(\omega)$ on adopte $\lg(\omega)$.

Si l'argument de la fonction

$$\Psi(\lg \omega) = \psi(\omega) = \arg G(i\omega) \quad (0 < \omega < \infty),$$

est $\lg(\omega)$, on l'appelle *réponse en phase logarithmique.*

La grandeur $20 \lg R(\omega)$ traduit le gain du système en décibels, unités adoptées en acoustique (1 décibel = 0,1 bel). Le gain qui augmente de 10^q fois la puissance du signal est égal à q bels. La puissance du signal est proportionnelle au carré de son amplitude. Soit, par exemple, $L(\omega_1) = 20 \lg R(\omega_1) = 40$ décibels. Dans ces conditions, $\lg R(\omega_1) = 2$, $R(\omega_1) = 10^2 = 100$. La puissance du signal s'est accrue de $100^2 = 10^4$ fois, ce qui correspond au gain d'amplification de 4 bels. Pour construire la courbe de la fonction $L(\omega)$, on porte en abscisses les valeurs de $\lg \omega$ et en ordonnées les valeurs de la fonction $L(\omega)$ en décibels.

Pour construire la courbe de la fonction $\psi(\omega)$, on porte en abscisses les valeurs de $\lg \omega$, et en ordonnées, les valeurs de $\psi(\omega)$ en radians.

Comme unité de longueur en abscisses on utilise les unités logarithmiques, octave et décade. On dit octave pour un segment de l'axe $\lg \omega$ compris entre ω_1 et $2\omega_1$. On voit aisément que la longueur de ce segment ne dépend pas de la valeur de ω_1

$$l_{\text{oct}} = \lg 2\omega_1 - \lg \omega_1 = \lg 2 + \lg \omega_1 - \lg \omega_1 = \lg 2.$$

On dit décade pour un segment de l'axe $\lg \omega$ compris entre ω_1 et $10\omega_1$. La longueur de la décade ne dépend pas de ω_1:

$$l_{\text{déc}} = \lg 10\omega_1 - \lg \omega_1 = \lg 10 + \lg \omega_1 - \lg \omega_1 = 1.$$

9. Détermination de la stabilité d'un système en boucle fermée d'après les courbes logarithmiques d'un système en boucle ouverte. La disposition relative des réponses en amplitude et en phase d'un système *en boucle ouverte* permet de juger de la stabilité d'un système *en boucle fermée.* Les figures 3.10 à 3.12 représentent les lieux de transfert et les réponses logarithmiques, en amplitude et en phase, des systèmes en boucle ouverte stables. Ces dessins montrent que les lieux de transfert (lieux de Nyquist) des figures 3.10 et 3.11 sont stables à l'état bouclé. Le système représenté sur la figure 3.12 est instable à l'état bouclé.

On voit sans peine que pour qu'un système en boucle fermée soit stable, s'il l'est en boucle ouverte, il faut que dans la plage des fréquences où la réponse en amplitude logarithmique est positive

(c'est-à-dire où $R(\omega) > 1$) les nombres de passages par la ligne $\psi = -\pi$ de la réponse en phase dans le sens positif (de haut en bas) et dans le sens négatif (de bas en haut) soient égaux.

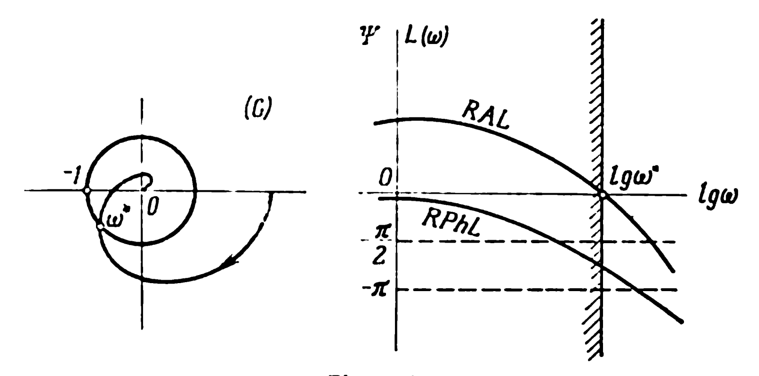


Fig. 3.10

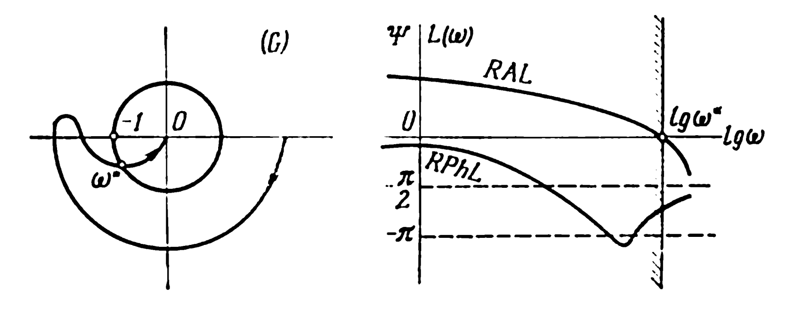


Fig. 3.11

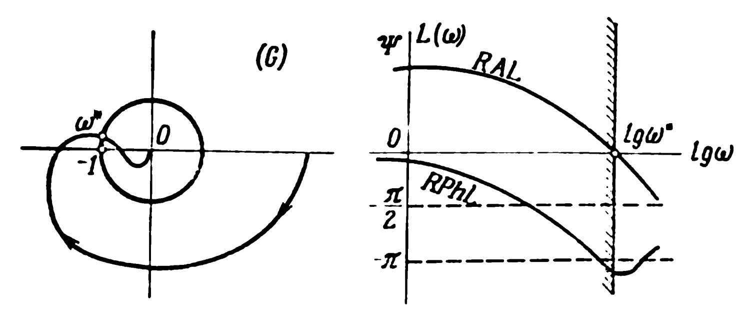


Fig. 3.12

Le critère que nous venons de décrire pour déterminer la stabilité d'un système en boucle fermée d'après les réponses logarithmiques

en boucle ouverte peut être généralisé au cas où la fonction de transfert du système en boucle ouverte $G(p)$ possède des pôles sur l'axe imaginaire et dans le demi-plan droit de la variable complexe p [69].

§ 4. Fonction de poids et réponse unitaire d'un système linéaire stationnaire

1. Système commandé monovariable. On appelle systèmes linéaires stationnaires les systèmes décrits par des équations différentielles linéaires à coefficients constants. Dans ce qui suit nous traiterons des systèmes *en boucle fermée*. Soit l'équation différentielle d'un système en boucle fermée

$$f(D)\,x = l(D)\,z(t). \tag{4.1}$$

Ici x est la fonction cherchée (signal de sortie), $z(t)$, la fonction de temps donnée (signal d'entrée) ; $f(D)$ et $l(D)$ les polynômes de l'opérateur de différentiation D :

$$\left.\begin{aligned} f(D) &= a_0D^n + a_1D^{n-1} + \ldots + a_{n-1}D + a_n, \\ l(D) &= l_0D^m + l_1D^{m-1} + \ldots + l_{m-1}D + l_m. \end{aligned}\right\} \tag{4.2}$$

Les coefficients des polynômes (2) sont supposés constants. De plus, on admet que $m < n$. Pour résoudre l'équation différentielle (1), on peut appliquer les méthodes du calcul opérationnel. Introduisons les notations

$$\zeta(p) \rightarrow z(t), \quad \xi(p) \rightarrow x(t). \tag{4.3}$$

La transformée de $\dfrac{d^kx}{dt^k}$ étant

$$\begin{aligned} D^kx \leftarrow p^k\xi(p) - p^kx(0) - p^{k-1}\dot{x}(0) - \ldots \\ \ldots - px^{(k-1)}(0) = p^k\xi(p) - p\frac{p^k - D_0^k}{p - D_0}\,x(0), \end{aligned} \tag{4.4}$$

les expressions différentielles linéaires $f(D)\,x(t)$ et $l(D)\,z(t)$ ont pour transformées

$$\left.\begin{aligned} f(D)\,x(t) &\leftarrow f(p)\,\xi(p) - pQ(p), \\ l(D)\,z(t) &\leftarrow l(p)\,\zeta(p) - pP(p), \end{aligned}\right\} \tag{4.5}$$

où

$$\left.\begin{aligned} Q(p) &= \frac{f(p) - f(D_0)}{p - D_0}\,x(0), \\ P(p) &= \frac{l(p) - l(D_0)}{p - D_0}\,z(0). \end{aligned}\right\} \tag{4.6}$$

L'opérateur $D_0^{\cdot}$ qui intervient dans les expressions (4) et (6) est déterminé par la relation

$$D_0^k x(0) = \left[\frac{d^k x}{dt^k}\right]_{t=0}. \tag{4.7}$$

D'après (5), à l'équation différentielle (1) correspond la transformation suivante

$$f(p)\,\xi(p) = p\,[Q(p) - P(p)] + l(p)\,\zeta(p). \tag{4.8}$$

On en tire que

$$\xi(p) = p\,\frac{Q(p) - P(p)}{f(p)} + \frac{l(p)}{f(p)}\,\zeta(p). \tag{4.9}$$

En adoptant pour la *fonction de transfert* du système (1) la notation

$$\Phi(D) = \frac{l(D)}{f(D)} \tag{4.10}$$

on peut mettre l'expression (9) sous la forme:

$$\xi(p) = p\,\frac{Q(p) - P(p)}{f(p)} + \Phi(p)\,\zeta(p). \tag{4.11}$$

Désignons maintenant par $M(t)$ et $N(t)$ les originaux des transformées suivantes:

$$p\,\frac{Q(p) - P(p)}{f(p)} \doteqdot M(t), \tag{4.12}$$

$$p\Phi(p) \doteqdot N(t). \tag{4.13}$$

Etant donné qu'en vertu de l'hypothèse ci-dessus suivant laquelle le degré du polynôme $l(p)$ est inférieur à celui du polynôme $f(p)$ les fonctions $[Q(p) - P(p)]/f(p)$ et $\Phi(p) = l(p)/f(p)$ sont des fractions régulières, les originaux $M(t)$ et $N(t)$ sont de la forme

$$M(t) = \sum_{\sigma=1}^{s} \frac{e^{\lambda_\sigma t}}{(q_\sigma - 1)!}\left[\left(t + \frac{\partial}{\partial p}\right)^{q_\sigma - 1} \frac{Q(p) - P(p)}{f_\sigma(p)}\right]_{p=\lambda_\sigma}, \tag{4.14}$$

$$N(t) = \sum_{\sigma=1}^{s} \frac{e^{\lambda_\sigma t}}{(q_\sigma - 1)!}\left[\left(t + \frac{\partial}{\partial p}\right)^{q_\sigma - 1} \frac{l(p)}{f_\sigma(p)}\right]_{p=\lambda_\sigma}. \tag{4.15}$$

Nous avons désigné ici par s le nombre de racines distinctes λ_σ de l'équation caractéristique

$$f(p) = 0, \tag{4.16}$$

par q_σ l'ordre de multiplicité de la racine λ_σ et par $f_\sigma(p)$ la fonction

$$f_\sigma(p) = \frac{f(p)}{(p - \lambda_\sigma)^{q_\sigma}}. \tag{4.17}$$

Il se peut que l'équation caractéristique (16) comporte des racines réelles et des racines complexes. Puisque les coefficients des polynômes (2) sont réels, les racines complexes de l'équation (16) forment des couples conjugués. Pour cette raison les fonctions $M(t)$ et $N(t)$ ne peuvent prendre que des valeurs réelles et dans le cas où l'équation (16) possède des racines complexes, les fonctions $M(t)$ et $N(t)$ peuvent être ramenées à une forme réelle explicite; mais nous n'allons pas nous attarder sur ce fait [17].

En vertu du théorème de multiplication des transformées

$$\frac{1}{p} p\Phi(p)\zeta(p) \risingdotseq \int_0^t N(t-u)\,z(u)\,du. \qquad (4.18)$$

Ainsi, d'après (11), (12) et (18), les conditions initiales $x(0)$, $\dot{x}(0), \ldots, x^{n-1}(0)$ étant données, la loi du mouvement du système décrit par l'équation (1)

$$x(t) = M(t) + \int_0^t N(t-u)\,z(u)\,du. \qquad (4.19)$$

On peut en déduire également la loi de variation dans le temps d'autres coordonnées de phase du système: $\dot{x}(t)$, $\ddot{x}(t), \ldots, x^{n-1}(t)$.

D'une façon analogue à l'exposé du § 1 (formule (1.31)), nous appellerons la fonction $N(t)$, définie ici par la relation opérationnelle (13), *fonction de poids* du système monovariable décrit par l'équation différentielle scalaire (1).

Considérons maintenant l'interprétation physique de la fonction $N(t)$. Soit

$$z(t) = \delta(t), \qquad (4.20)$$

c'est-à-dire on applique à l'entrée du système la fonction impulsion de Dirac. L'équation (1) se met alors sous la forme

$$f(D)\,x = l(D)\,\delta(t). \qquad (4.21)$$

Puisqu'à l'original

$$l(D)\,\delta(t) = l_0\delta^{(m)}(t) + l_1\delta^{(m-1)}(t) + \ldots + l_{m-1}\dot{\delta}(t) + l_m\delta(t)$$

correspond la transformée

$$l_0p^{m+1} + l_1p^m + \ldots + l_{m-1}p^2 + l_mp = pl(p),$$

c'est-à-dire

$$l(D)\,\delta(t) \risingdotseq pl(p), \qquad (4.22)$$

alors en tenant compte qu'en vertu de (5)

$$f(D)\,x(t) \risingdotseq f(p)\,\xi(p) - pQ(p),$$

on trouve qu'à l'équation différentielle (21) correspond la transformation

$$f(p)\,\xi(p) = pQ(p) + pl(p). \tag{4.23}$$

L'équation (23) entraîne que

$$\xi(p) = p\frac{Q(p)}{f(p)} + p\Phi(p), \tag{4.24}$$

du fait que d'après (10) $\Phi(p) = l(p)/f(p)$. Dans les conditions initiales nulles

$$x(0) = 0, \quad \dot{x}(0) = 0, \ \ldots, \ x^{(n-1)}(0) = 0 \tag{4.25}$$

on a d'après (4) et (6)

$$Q(p) = 0.$$

Ainsi, dans les conditions initiales nulles, l'expression (24) s'écrit

$$\xi(p) = p\Phi(p). \tag{4.26}$$

Compte tenu de la relation opérationnelle (13), on a

$$x(t) = N(t). \tag{4.27}$$

Ainsi, la fonction de poids $N(t)$ est une loi du mouvement du système dans les conditions initiales nulles sous l'effet de la fonction impulsion de Dirac $\delta(t)$. Notons maintenant qu'en vertu du théorème de la construction de l'original d'après sa transformée, la relation opérationnelle (13), qui a servi à déterminer la fonction $N(t)$, ne définit cette fonction que pour $t \geqslant 0$. Par suite l'expression (15) qui définit la fonction $N(t)$ n'a lieu que pour $t \geqslant 0$. Comme d'après (25) pour $t < 0$ le système était en équilibre, nous devons adopter en vertu de (27) que

$$N(t) = 0 \text{ pour } t < 0. \tag{4.28}$$

Pour trouver une autre caractéristique dynamique importante d'un système, dite *réponse unitaire*, considérons le cas où le signal d'entrée est une fonction échelon unitaire

$$z(t) = 1(t). \tag{4.29}$$

L'équation différentielle initiale (1) s'écrit

$$f(D)\,x = l(D)\,1(t). \tag{4.30}$$

A l'original

$$l(D)\,1(t) = l_0\delta^{(m-1)}(t) + l_1\delta^{(m-2)}(t) + \ldots + l_{m-1}\delta(t) + l_m 1(t)$$

correspond la transformée

$$l_0p^m + l_1p^{m-1} + \ldots + l_{m-1}p + l_m = l(p), \tag{4.31}$$

c'est-à-dire

$$l(D)\,1(t) \risingdotseq l(p). \tag{4.32}$$

Les relations (32) et (5) entraînent qu'à l'équation différentielle (30) correspond la transformation

$$f(p)\,\xi(p) = pQ(p) + l(p), \tag{4.33}$$

d'où

$$\xi(p) = p\frac{Q(p)}{f(p)} + \Phi(p). \tag{4.34}$$

Dans les conditions initiales nulles (25), $Q(p) = 0$ et l'expression (34) se met sous la forme

$$\xi(p) = \Phi(p). \tag{4.35}$$

Désignons par $A(t)$ l'original dont la transformée est la fonction $\Phi(p)$:

$$\Phi(p) \to A(t). \tag{4.36}$$

La relation (35) entraîne

$$x(t) = A(t). \tag{4.37}$$

De cette façon, la fonction $A(t)$ définie par la relation opérationnelle (36) est la loi du mouvement du système dans les conditions initiales nulles sous l'effet de la fonction échelon unitaire $1(t)$.

La fonction $A(t)$ s'appelle *réponse unitaire* du système.

Cherchons maintenant la relation entre la réponse unitaire et la fonction de poids. Remarquons que puisque le degré du polynôme $l(p)$ est inférieur au degré du polynôme $f(p)$, il vient

$$\lim_{p\to\infty} \Phi(p) = \lim_{p\to\infty} \frac{l(p)}{f(p)} = 0. \tag{4.38}$$

D'après le théorème sur la valeur initiale de l'original

$$A(0) = \lim_{p\to\infty} \Phi(p)$$

et d'après (38)

$$A(0) = 0.$$

La transformée de la dérivée de la fonction $A(t)$

$$\dot{A}(t) \leftarrow p[\Phi(p) - A(0)] = p\Phi(p) \tag{4.39}$$

se confond avec la transformée (13) de la fonction $N(t)$. Ainsi, la fonction de poids $N(t)$ et la réponse unitaire $A(t)$ sont liées par la relation

$$\dot{A}(t) = N(t). \tag{4.40}$$

Cherchons maintenant l'expression explicite de la fonction $A(t)$. A cet effet mettons la transformée de la fonction $A(t)$ sous la forme

d'une somme de fractions élémentaires. Puisque

$$\frac{l(p)}{pf(p)}=\frac{l(0)}{f(0)}\frac{1}{p}+$$

$$+\sum_{\sigma=1}^{s}\sum_{\rho=0}^{q_\sigma-1}\frac{1}{(q_\sigma-\rho-1)!}\left[\frac{\partial^{q_\sigma-\rho-1}}{\partial p^{q_\sigma-\rho-1}}\frac{l(p)}{pf_\sigma(p)}\right]_{p=\lambda_\sigma}\frac{1}{(p-\lambda_\sigma)^{\rho+1}}, \quad (4.41)$$

la décomposition de la fonction $\Phi(p)$ en une somme de fractions élémentaires s'écrit

$$\Phi(p)=\frac{l(p)}{f(p)}=\frac{l(0)}{f(0)}+$$

$$+\sum_{\sigma=1}^{s}\sum_{\rho=0}^{q_\sigma-1}\frac{1}{(q_\sigma-\rho-1)!}\left[\frac{\partial^{q_\sigma-\rho-1}}{\partial p^{q_\sigma-\rho-1}}\frac{l(p)}{pf_\sigma(p)}\right]_{p=\lambda_\sigma}\frac{p}{(p-\lambda_\sigma)^{\rho+1}}. \quad (4.42)$$

Etant donné que

$$\frac{p}{(p-\lambda_\sigma)^{\rho+1}}\;\dot{\to}\;\frac{t^\rho e^{\lambda_\sigma t}}{\rho!}, \quad (4.43)$$

on a l'original de la transformée (42)

$$A(t)=\frac{l(0)}{f(0)}+\sum_{\sigma=1}^{s}\sum_{\rho=0}^{q_\sigma-1}\frac{t^\rho e^{\lambda_\sigma t}}{\rho!\,(q_\sigma-\rho-1)!}\left[\frac{\partial^{q_\sigma-\rho-1}}{\partial p^{q_\sigma-\rho-1}}\frac{l(p)}{pf_\sigma(p)}\right]_{p=\lambda_\sigma}. \quad (4.44)$$

On vérifie aisément que

$$\sum_{\rho=0}^{q_\sigma-1}\frac{t^\rho}{\rho!\,(q_\sigma-\rho-1)!}\frac{\partial^{q_\sigma-\rho-1}}{\partial p^{q_\sigma-\rho-1}}=\frac{1}{(q_\sigma-1)!}\left(t+\frac{\partial}{\partial p}\right)^{q_\sigma-1}, \quad (4.45)$$

et par suite l'expression (44) de la réponse unitaire peut se mettre sous la forme

$$A(t)=\frac{l(0)}{f(0)}+\sum_{\sigma=1}^{s}\frac{e^{\lambda_\sigma t}}{(q_\sigma-1)!}\left[\left(t+\frac{\partial}{\partial p}\right)^{q_\sigma-1}\frac{l(p)}{pf_\sigma(p)}\right]_{p=\lambda_\sigma}. \quad (4.46)$$

D'après ce qui vient d'être dit, on peut conclure que pour un système commandé, la fonction de transfert $\Phi(D)$, la fonction de poids $N(t)$ et la réponse unitaire $A(t)$ sont les caractéristiques dynamiques principales, équivalentes entre elles.

2. Système commandé monovariable à fonction de transfert sous forme de fraction irrégulière. Considérons maintenant le système

décrit par l'équation différentielle scalaire

$$f(D)\,x(t) = r(D)\,z(t), \tag{4.47}$$

où

$$\begin{aligned} f(D) &= a_0D^n + a_1D^{n-1} + \ldots + a_{n-1}D + a_n, \\ r(D) &= b_0D^q + b_1D^{q-1} + \ldots + b_{q-1}D + b_q, \end{aligned} \tag{4.48}$$

et de plus, $q > n$.

La fonction de transfert du système

$$\Phi(D) = \frac{r(D)}{f(D)} \tag{4.49}$$

peut être mise sous la forme

$$\Phi(D) = S(D) + \nu(D), \tag{4.50}$$

où $S(D)$ est la partie entière de la fraction rationnelle $r(D)/f(D)$ (c'est-à-dire un certain polynôme en D), et $\nu(D)$ la fraction régulière qui s'obtient en séparant de $r(D)/f(D)$ la partie entière.

En utilisant les mêmes notations que dans (5)

$$\xi(p) \div\!\!\!\rightarrow x(t), \quad \zeta(p) \div\!\!\!\rightarrow z(t),$$

on a

$$\left.\begin{aligned} f(D)\,x(t) &\leftarrow\!\!\!\div f(p)\,\xi(p) - pQ(p), \\ r(D)\,z(t) &\leftarrow\!\!\!\div r(p)\,\zeta(p) - pL(p), \end{aligned}\right\} \tag{4.51}$$

où

$$\left.\begin{aligned} Q(p) &= \frac{f(p) - f(D_0)}{p - D_0}\,x(0), \\ L(p) &= \frac{r(p) - r(D_0)}{p - D_0}\,z(0). \end{aligned}\right\} \tag{4.52}$$

En vertu de (51) on a la transformation

$$f(p)\,\xi(p) = p\,[Q(p) - L(p)] + r(p)\,\zeta(p). \tag{4.53}$$

On en tire

$$\xi(p) = p\,\frac{Q(p) - L(p)}{f(p)} + \frac{r(p)}{f(p)}\,\zeta(p). \tag{4.54}$$

Il est facile de voir que la fonction $\frac{1}{f(p)}[Q(p) - L(p)]$ est également une fraction irrégulière. En effet, d'après (52), on a

$$\frac{Q(p) - L(p)}{f(p)} = \frac{f(p)\,x(0) - f(D_0)\,x(0) - r(p)\,z(0) + r(D_0)\,z(0)}{f(p)\,(p - D_0)}. \tag{4.55}$$

Puisque d'après (49) et (50)

$$r(p) = f(p)\,S(p) + f(p)\,\nu(p), \tag{4.56}$$

l'expression (55) peut alors être ramenée à la forme

$$\frac{Q(p)-L(p)}{f(p)}=-\frac{S(p)-S(D_0)}{p-D_0}z(0)+\frac{1}{f(p)}\left\{\frac{f(p)-f(D_0)}{p-D_0}[x(0)-\right.$$
$$\left.-S(D_0)z(0)]-\frac{f(p)\nu(p)-f(D_0)\nu(D_0)}{p-D_0}z(0)\right\}. \qquad (4.57)$$

Le premier terme du second membre de (57) est une fonction entière (polynôme en p). Le deuxième terme est une fonction rationnelle de p, qui est de plus une fraction régulière. Comme $f(p)\,\nu(p)$ est un polynôme en p, on élimine le dénominateur $p - D_0$ dans le deuxième terme du second membre de l'expression (57).

L'expression (57) pourra alors s'écrire:

$$\frac{Q(p)-L(p)}{f(p)}=-R(p)+\mu(p), \qquad (4.58)$$

avec

$$R(p)=\frac{S(p)-S(D_0)}{p-D_0}z(0), \qquad (4.59)$$

et $\mu(p)$ est une fraction régulière déterminée par le deuxième terme du second membre de (57).

La fonction $\mu(p)$ peut être mise sous la forme d'une somme des fractions élémentaires [17]:

$$\mu(p)=\sum_{\sigma=1}^{s}\sum_{\rho=0}^{q_\sigma-1}\frac{1}{(q_\sigma-\rho-1)!}\left[\frac{\partial^{q_\sigma-\rho-1}}{\partial p^{q_\sigma-\rho-1}}\times\right.$$
$$\left.\times\frac{Q(p)-L(p)}{f_\sigma(p)}\right]_{p=\lambda_\sigma}\frac{1}{(p-\lambda_\sigma)^{\rho+1}}. \qquad (4.60)$$

L'original dont la transformée est la fonction $p\mu(p)$ est noté $M^*(t)$:

$$p\mu(p) \risingdotseq M^*(t). \qquad (4.61)$$

Compte tenu de la relation (45), mettons la fonction $M^*(t)$ sous la forme

$$M^*(t)=\sum_{\sigma=1}^{s}\frac{e^{\lambda_\sigma t}}{(q_\sigma-1)!}\left[\left(t+\frac{\partial}{\partial p}\right)^{q_\sigma-1}\frac{Q(p)-L(p)}{f_\sigma(p)}\right]_{p=\lambda_\sigma}. \qquad (4.62)$$

Si l'on applique (58) et (50), l'expression (54) qui définit la fonction $\xi(p)$ peut également s'écrire

$$\xi(p) = p\mu(p) + S(p)\,\zeta(p) - pR(p) + \nu(p)\,\zeta(p). \qquad (4.63)$$

En tenant compte de la relation (59) on trouve

$$S(p)\zeta(p) - pR(p) \risingdotseq S(D)\,z(t). \qquad (4.64)$$

Désignons l'original de la transformée $p\nu(p)$ par $N^*(t)$:

$$p\nu(p) \doteqdot N^*(t). \tag{4.65}$$

D'une façon analogue à (62), la fonction $N^*(t)$ est de la forme:

$$N^*(t) = \sum_{\sigma=1}^{s} \frac{e^{\lambda_\sigma t}}{(q_\sigma - 1)!} \left[\left(t + \frac{\partial}{\partial p}\right)^{q_\sigma - 1} \frac{r(p)}{f_\sigma(p)}\right]_{p=\lambda_\sigma}. \tag{4.66}$$

Comme

$$\frac{1}{p} p\nu(p)\zeta(p) \doteqdot \int_0^t N^*(t-u)\, z(u)\, du, \tag{4.67}$$

en vertu de (63) la loi du mouvement du système régi par l'équation (47) s'écrit

$$x(t) = M^*(t) + S(D) z(t) + \int_0^t N^*(t-u)\, z(u)\, du. \tag{4.68}$$

Examinons maintenant l'équation différentielle scalaire

$$f(D)\, x(t) = r(D)\, \delta(t), \tag{4.69}$$

où $\delta(t)$ est la fonction impulsion de Dirac.

D'une façon analogue à (22)

$$r(D)\, \delta(t) \doteqdot pr(p). \tag{4.70}$$

Puisque d'après (51)

$$f(D)\, x(t) \doteqdot f(p)\, \xi(p) - pQ(p),$$

à l'équation différentielle (69) correspond la transformation

$$f(p)\, \xi(p) = pQ(p) + pr(p), \tag{4.71}$$

d'où

$$\xi(p) = p\frac{Q(p)}{f(p)} + p\frac{r(p)}{f(p)}. \tag{4.72}$$

Conformément à (49) et (50), l'expression (72) peut s'écrire

$$\xi(p) = p\frac{Q(p)}{f(p)} + pS(p) + p\nu(p). \tag{4.73}$$

Dans les conditions initiales nulles

$$x(0) = 0,\ \dot{x}(0) = 0,\ \ldots,\ \overset{n-1}{x}(0) = 0 \tag{4.74}$$

la fonction $Q(p)$ s'annule en vertu de (4) et (6)

$$Q(p) = 0,$$

et l'expression (73) devient

$$\xi(p) = p\Phi(p), \quad (4.75)$$

où, d'après (50),

$$\Phi(p) = S(p) + \nu(p).$$

$S(p)$ étant un polynôme en p, il vient

$$pS(p) \doteqdot S(D)\,\delta(t). \quad (4.76)$$

Si l'on tient compte encore de la relation (65)

$$p\nu(p) \doteqdot N^*(t),$$

on trouve qu'à la transformée $p\Phi(p)$ correspond l'original

$$p\Phi(p) \doteqdot K(t), \quad (4.77)$$

avec

$$K(t) = S(D)\,\delta(t) + N^*(t). \quad (4.78)$$

En vertu de (75) et (77)

$$x(t) = K(t). \quad (4.79)$$

Ainsi, la fonction $K(t)$ est la loi du mouvement réalisé par un système commandé dans les conditions initiales nulles sous l'effet de la fonction impulsion de Dirac $\delta(t)$.

La fonction $K(t)$ déterminée par la relation opérationnelle (77) peut s'appeller *fonction de poids* du système commandé monovariable décrit par l'équation différentielle scalaire (47).

De même qu'au point 1, l'expression (78) ne détermine la fonction $K(t)$ que pour $t \geqslant 0$. Etant donné qu'avec $t < 0$ le système était en équilibre, en vertu de (79) nous devons adopter que

$$K(t) = 0 \text{ pour } t < 0. \quad (4.80)$$

Considérons maintenant l'équation

$$f(D)\,x(t) = r(D)\,1(t). \quad (4.81)$$

D'une façon analogue à (32)

$$r(D)\,1(t) \risingdotseq r(p).$$

En tenant compte de la relation (51), on obtient la transformation

$$f(p)\,\xi(p) = pQ(p) + r(p), \quad (4.82)$$

qui correspond à l'équation différentielle (81), d'où

$$\xi(p) = p\frac{Q(p)}{f(p)} + \frac{r(p)}{f(p)}, \quad (4.83)$$

ou, en vertu de (49) et (50),

$$\xi(p) = p\frac{Q(p)}{f(p)} + S(p) + \nu(p). \quad (4.84)$$

Sous les conditions initiales nulles (74)

$$x(0)=0,\quad \dot{x}(0)=0,\ \ldots,\ \overset{n-1}{x}(0)=0$$

la fonction $Q(p) = 0$ et l'expression (84) se met sous la forme

$$\xi(p) = \Phi(p), \tag{4.85}$$

où d'après (50),

$$\Phi(p) = S(p) + \nu(p).$$

La fonction $S(p)$ est un polynôme en p, et par conséquent,

$$S(p) \doteqdot S(D)\,1(t). \tag{4.86}$$

Désignons par $A^*(t)$ l'original dont la transformée est la fonction $\nu(p)$:

$$\nu(p) \doteqdot A^*(t). \tag{4.87}$$

En vertu de (46) et (66), la fonction $A^*(t)$ s'écrit

$$A^*(t)=\frac{r(0)}{f(0)}+\sum_{\sigma=1}^{s}\frac{e^{\lambda_\sigma t}}{(q_\sigma-1)!}\left[\left(t+\frac{\partial}{\partial p}\right)^{q_\sigma-1}\frac{r(p)}{pf_\sigma(p)}\right]_{p=\lambda_\sigma}. \tag{4.88}$$

Ainsi, à la transformée $\Phi(p)$ correspond l'original

$$\Phi(p) \doteqdot U(t), \tag{4.89}$$

avec

$$U(t) = S(D)\,1(t) + A^*(t). \tag{4.90}$$

Conformément à (85) et (89),

$$x(t) = U(t), \tag{4.91}$$

et, par suite, la fonction $U(t)$ est la loi du mouvement du système commandé dans les conditions initiales nulles sous l'effet de l'échelon unitaire $1(t)$.

La fonction $U(t)$ déterminée par la relation opérationnelle (89), peut s'appeler *réponse unitaire* d'un système monovariable décrit par l'équation différentielle scalaire (47).

Notons que puisque $\nu(p)$ est une fraction régulière, $\lim_{p\to\infty}\nu(p) = 0$, et, par conséquent, $A^*(0) = 0$. C'est pourquoi

$$\dot{A}^*(t) \doteqdot p\,[\nu(p) - A^*(0)] = p\nu(p). \tag{4.92}$$

Si l'on retient encore que

$$\frac{d}{dt}[S(D)\,1(t)] = S(D)\,\delta(t) \doteqdot pS(p), \tag{4.93}$$

on a

$$\dot{U}(t) \doteqdot pS(p) + p\nu(p) = p\Phi(p), \tag{4.94}$$

ce qui coïncide avec la transformée (77) de la fonction $K(t)$. On en déduit une liaison analogue à la relation (40) entre la fonction de poids $K(t)$ et la réponse unitaire $U(t)$:

$$\dot{U}(t) = K(t). \tag{4.95}$$

Exemple. Examinons à titre d'exemple le système décrit par l'équation différentielle

$$(D+1)\,x = (D^2+D+1)\,z(t). \tag{4.96}$$

Ici

$$f(D) = D+1, \qquad r(D) = D^2+D+1,$$

$$\Phi(D) = \frac{r(D)}{f(D)} = S(D) + \nu(D) = D + \frac{1}{D+1},$$

$$S(D) = D, \qquad \nu(D) = \frac{1}{D+1},$$

$$Q(p) = \frac{f(p)-f(D_0)}{p-D_0}\,x(0) = x(0),$$

$$L(p) = \frac{r(p)-r(D_0)}{p-D_0}\,z(0) = \frac{p^2+p-D_0^2-D_0}{p-D_0}\,z(0) = (p+D_0+1)\,z(0),$$

$$\frac{Q(p)-L(p)}{f(p)} = -R(p)+\mu(p) = \frac{x(0)-(p+D_0+1)\,z(0)}{p+1} = -z(0) + \frac{x(0)-\dot{z}(0)}{p+1},$$

$$R(p) = z(0), \qquad \mu(p) = \frac{x(0)-\dot{z}(0)}{p+1}, \qquad p\mu(p) \risingdotseq M^*(t),$$

$$M^*(t) = [x(0)-\dot{z}(0)]\,e^{-t}, \qquad p\nu(p) \risingdotseq N^*(t), \; N^*(t) = e^{-t}.$$

D'après l'expression (68), la solution de l'équation (96) est

$$x(t) = [x(0)-\dot{z}(0)]\,e^{-t} + \dot{z}(t) + \int_0^t e^{-(t-u)}z(u)\,du. \tag{4.97}$$

En vertu de (78), la fonction de poids du système $K(t)$ s'écrit

$$K(t) = S(D)\,\delta(t) + N^*(t).$$

Pour le système décrit par l'équation différentielle (96) la fonction $K(t)$ se met sous la forme

$$K(t) = \dot{\delta}(t) + 1(t)\,e^{-t}. \tag{4.98}$$

Le facteur $1(t)$ est introduit dans l'expression (98) pour expliciter le fait qu'avec $t < 0$, $K(t) = 0$.

3. Systèmes commandés multivariables. a) Loi du mouvement d'un système multivariable. Considérons maintenant un système en boucle fermée décrit par l'équation différentielle vectorielle

$$f(D)\,x(t) = r(D)\,z(t). \tag{4.99}$$

Ici $x(t)$ est le vecteur de dimension n dont les éléments sont les fonctions recherchées, et $z(t)$, le vecteur de dimension m dont les éléments sont les fonctions du temps données:

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \dots \\ x_n(t) \end{bmatrix}, \quad z(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \\ \dots \\ z_m(t) \end{bmatrix}. \tag{4.100}$$

Les matirices $f(D)$ et $r(D)$ sont des matrices polynomiales dont la variable est l'opérateur de différentiation $D = d/dt$:

$$f(D) = \begin{bmatrix} f_{11}(D) \dots f_{1n}(D) \\ \dots\dots\dots \\ f_{n1}(D) \dots f_{nn}(D) \end{bmatrix},$$

$$r(D) = \begin{bmatrix} r_{11}(D) \dots r_{1m}(D) \\ \dots\dots\dots \\ r_{n1}(D) \dots r_{nm}(D) \end{bmatrix}. \tag{4.101}$$

Les coefficients des polynômes $f_{jk}(D)$ et $r_{j\mu}(D)$ sont supposés constants.

L'équation vectorielle (99) est équivalente au système des équations scalaires

$$\sum_{k=1}^{n} f_{jk}(D)\, x_k(t) = \sum_{\mu=1}^{m} r_{j\mu}(D)\, z_\mu(t) \quad (j = 1, \dots, n). \tag{4.102}$$

Désignons par ν_k l'ordre de la dérivée supérieure par rapport à x_k ($k = 1, \dots, n$) qu'on trouve dans les équations (102). Le rang ρ du système (102) est

$$\rho = \nu_1 + \nu_2 + \dots + \nu_n. \tag{4.103}$$

Supposons que $\nu_k \geqslant 1$ ($k = 1, \dots, n$) et que le déterminant constitué par les coefficients affectés aux $x_k^{(\nu_k)}$ ($k = 1, \dots, n$) de l'équation (102) est différent du zéro. Sous cette condition, le degré du déterminant $\Delta(D)$ de la matrice $f(D)$ est égal au rang du système ρ.

D'après ces hypothèses, les valeurs initiales

$$x_k(0), \ \dot{x}_k(0), \ \dots, \ \overset{\nu_k - 1}{x_k}(0) \quad (k = 1, \dots, n)$$

ne sont liées entre elles par aucune relation et peuvent être données arbitrairement.

Si l'on désigne par $F(D)$ la matrice adjointe de la matrice $f(D)$, l'équation (99) peut se mettre sous la forme

$$x(t) = \Phi(D)\, z(t), \tag{4.104}$$

où

$$\Phi(D) = \frac{F(D)\, r(D)}{\Delta(D)}. \tag{4.105}$$

La fonction $\Phi(D)$ est la *matrice de transfert* du système décrit par l'équation (99).

Supposons pour la généralité que la fonction $\Phi(D)$ est une fraction irrégulière; désignons par $S(D)$ sa partie entière et par $\nu(D)$ la fraction régulière qui s'obtient en séparant cette partie entière de $\Phi(D)$. Ainsi

$$\Phi(D) = \frac{F(D)\, r(D)}{\Delta(D)} = S(D) + \nu(D). \tag{4.106}$$

Dans ce qui suit nous aurons besoin d'une relation qui s'obtient de la façon suivante. En multipliant à gauche le premier et le second membre de (106) par la matrice $f(D)$ et en tenant compte du fait que $f(D)\, F(D) = E\Delta(D)$, où E est la matrice unité, on a

$$r(D) = f(D)\, S(D) + f(D)\, \nu(D). \tag{4.107}$$

Comme $r(D)$ et $f(D)\, S(D)$ sont des fonctions entières (polynômes en D), (107) entraîne que $f(D)\, \nu(D)$ est une fonction entière (polynôme en D).

Cherchons maintenant la solution de l'équation (99). Introduisons les notations

$$\xi(p) \to x(t), \quad \zeta(p) \to z(t). \tag{4.108}$$

En vertu de la relation opérationnelle (4), les transformées des expressions différentielles linéaires vectorielles $f(D)x(t)$ et $r(D)\, z(t)$ sont

$$\left.\begin{aligned} f(D)\, x(t) &\leftarrow f(p)\, \xi(p) - pQ(p), \\ r(D)\, z(t) &\leftarrow r(p)\, \zeta(p) - pL(p), \end{aligned}\right\} \tag{4.109}$$

où $Q(p)$ et $L(p)$ sont des vecteurs qui s'écrivent

$$\left.\begin{aligned} Q(p) &= \frac{f(p) - f(D_0)}{p - D_0}\, x(0), \\ L(p) &= \frac{r(p) - r(D_0)}{p - D_0}\, z(0). \end{aligned}\right\} \tag{4.110}$$

La transformation qui correspond à l'équation différentielle (99) est la suivante:

$$f(p)\, \xi(p) = p\,[Q(p) - L(p)] + r(p)\, \zeta(p), \tag{4.111}$$

d'où l'on tire

$$\xi(p) = p\, \frac{F(p)\,[Q(p) - L(p)]}{\Delta(p)} + \frac{F(p)\, r(p)}{\Delta(p)}\, \zeta(p). \tag{4.112}$$

Transformons maintenant le premier terme du second membre de la relation (112). En remplaçant $Q(p)$ et $L(p)$ par leurs expressions (110), on obtient après des transformations assez simples

$$\frac{F(p)\,[Q(p)-L(p)]}{\Delta(p)}=-\frac{S(p)-S(D_0)}{p-D_0}z(0)+$$
$$+\frac{F(p)}{\Delta(p)}\left\{\frac{f(p)-f(D_0)}{p-D_0}x(0)-\frac{[F^{-1}(p)\,\Delta(p)-F^{-1}(D_0)\,\Delta(D_0)]\,S(D_0)}{p-D_0}z(0)-\right.$$
$$\left.-\frac{F^{-1}(p)\,\Delta(p)\,\nu(p)-F^{-1}(D_0)\,\Delta(D_0)\,\nu(D_0)}{p-D_0}z(0)\right\}. \qquad (4.113)$$

En retenant que $F(p)\,f(p)=E\Delta(p)$ et, par suite,

$$F^{-1}(p)\,\Delta(p)=f(p),$$

l'expression (113) peut être mise sous la forme:

$$\frac{F(p)\,[Q(p)-L(p)]}{\Delta(p)}=-\frac{S(p)-S(D_0)}{p-D_0}z(0)+$$
$$+\frac{F(p)}{\Delta(p)}\left\{\frac{f(p)-f(D_0)}{p-D_0}[x(0)-S(D_0)\,z(0)]-\right.$$
$$\left.-\frac{f(p)\,\nu(p)-f(D_0)\,\nu(D_0)}{p-D_0}z(0)\right\}. \qquad (4.114)$$

Le premier terme du second membre de l'expression (114) est une fonction entière (polynôme matriciel en p), alors que le deuxième terme est une fonction rationnelle de p, qui est en plus une fraction régulière. Constatons que puisque $f(p)\,\nu(p)$ est une fonction entière (polynôme matriciel en p), comme on l'a vu plus haut (107), le dénominateur $p-D_0$ du deuxième terme du second membre de l'expression (114) se simplifie par $p-D_0$.

L'expression (114) peut donc s'écrire

$$\frac{F(p)\,[Q(p)-L(p)]}{\Delta(p)}=-R(p)+\mu(p), \qquad (4.115)$$

avec

$$R(p)=\frac{S(p)-S(D_0)}{p-D_0}z(0), \qquad (4.116)$$

alors que $\mu(p)$ est une fraction régulière déterminée par le deuxième terme du second membre de l'expression (114). Introduisons la notation $M(t)$ pour l'original dont la transformée est la fonction $p\mu(p)$:

$$p\mu(p) \div M(t). \qquad (4.117)$$

Dans ce qui suit nous aurons besoin encore de la transformation suivante. Le déterminant $\Delta(p)$ de la matrice $f(p)$ peut être mis sous

la forme

$$\Delta(p) = k_0 \mathscr{D}_n(p), \tag{4.118}$$

où

$$\mathscr{D}_n(p) = (p-\lambda_1)^{q_1} \ldots (p-\lambda_s)^{q_s} = \prod_{\sigma=1}^{s} (p-\lambda_\sigma)^{q_\sigma}. \tag{4.119}$$

Ainsi, λ_σ sont les zéros du polynôme $\Delta(p)$, et q_σ leurs ordres de multiplicité.

Soit $\mathscr{D}^*_{n-1}(p)$ le plus grand diviseur commun des mineurs d'ordre $(n-1)$ de la matrice $f(p)$, et α_0 le coefficient affecté au plus grand degré de p du polynôme $\mathscr{D}^*_{n-1}(p)$. Désignons par $\mathscr{D}_{n-1}(p)$ le polynôme suivant:

$$\mathscr{D}_{n-1}(p) = \frac{1}{\alpha_0} \mathscr{D}^*_{n-1}(p).$$

Ce polynôme est un diviseur déterminant d'ordre $(n-1)$ de la matrice polynomiale $f(p)$.

On sait [17, 21] que la fonction

$$\mathscr{L}_n(p) = \frac{\mathscr{D}_n(p)}{\mathscr{D}_{n-1}(p)} \tag{4.120}$$

est un polynôme de la forme

$$\mathscr{L}_n(p) = (p-\lambda_1)^{m_1} \ldots (p-\lambda_s)^{m_s} = \prod_{\sigma=1}^{s} (p-\lambda_\sigma)^{m_\sigma},$$

en outre

$$1 \leqslant m_\sigma \leqslant q_\sigma. \tag{4.121}$$

De cette façon les zéros du polynôme $\mathscr{L}_n(p)$ coïncident avec les zéros du polynôme $\Delta(p)$, alors que les multiplicités des zéros du polynôme $\mathscr{L}_n(p)$ vérifient la condition (121). Le polynôme $\mathscr{L}_n(p)$ est un facteur invariant supérieur de la matrice $f(p)$, alors que les binômes $(p-\lambda_\sigma)^{m_\sigma}$ sont des diviseurs élémentaires supérieurs de cette matrice.

Avec (118) et (120), on a

$$\Delta(p) = k_0 \mathscr{D}_{n-1}(p) \mathscr{L}_n(p).$$

Les éléments de la matrice $F(p)$ étant des cofacteurs des éléments du déterminant de la matrice $f(p)$, tous les éléments $F_{jk}(p)$ de la matrice $F(p)$ sont divisibles par le polynôme $\mathscr{D}_{n-1}(p)$ et pour la matrice $F(p)$ on peut écrire

$$F(p) = \mathscr{F}(p)\, \mathscr{D}_{n-1}(p).$$

La matrice $\mathscr{F}(p)$ est donc une matrice polynomiale.

Des relations obtenues ici il ressort que

$$\frac{F(p)}{\Delta(p)} = \frac{\mathcal{F}(p)}{\Lambda(p)}, \tag{4.122}$$

avec

$$\Lambda(p) = k_0 \mathcal{L}_n(p),$$

c'est-à-dire

$$\Lambda(p) = k_0 (p-\lambda_1)^{m_1} \ldots (p-\lambda_s)^{m_s} = k_0 \prod_{\sigma=1}^{s} (p-\lambda_\sigma)^{m_\sigma}. \tag{4.123}$$

Cherchons maintenant la fonction $M(t)$ qui en vertu de (117) est l'original de la transformée $p\mu(p)$. A cet effet, en vertu de (122), remplaçons dans l'expression (115) $F(p)/\Delta(p)$ par $\mathcal{F}(p)/\Lambda(p)$. La fonction $M(t)$ devient alors

$$M(t) = \sum_{\sigma=1}^{s} \frac{e^{\lambda_\sigma t}}{(m_\sigma - 1)!} \Big[\Big(t + \frac{\partial}{\partial p}\Big)^{m_\sigma - 1} \times \\ \times \frac{\mathcal{F}(p)[Q(p) - L(p)]}{\Lambda_\sigma(p)} \Big]_{p=\lambda_\sigma}. \tag{4.124}$$

Ici s est le nombre de zéros distincts λ_σ du polynôme $\Lambda(p)$ qui ne se confondent pas ; m_σ, la multiplicité du zéro λ_σ, et $\Lambda_\sigma(p)$, la fonction

$$\Lambda_\sigma(p) = \frac{\Lambda(p)}{(p-\lambda_\sigma)^{m_\sigma}}. \tag{4.125}$$

La fonction $M(t)$ est un vecteur de dimension n.

En appliquant (115) et (106) l'expression (112) qui détermine la fonction $\xi(p)$ peut se mettre sous la forme

$$\xi(p) = p\mu(p) + S(p)\zeta(p) - pR(p) + \nu(p)\zeta(p). \tag{4.126}$$

Compte tenu de la relation (116), on trouve que

$$S(p)\zeta(p) - pR(p) \doteq S(D)z(t). \tag{4.127}$$

Il résulte de l'exposé précédent que $S(D)$ est une matrice polynomiale $n \times m$ dont la variable est l'opérateur de différentiation $D = d/dt$. Désignons par $N(t)$ l'original de la transformée $p\nu(p)$:

$$p\nu(p) \doteq N(t). \tag{4.128}$$

D'une façon analogue à (124), la fonction $N(t)$ s'écrit

$$N(t) = \sum_{\sigma=1}^{s} \frac{e^{\lambda_\sigma t}}{(m_\sigma - 1)!} \Big[\Big(t + \frac{\partial}{\partial p}\Big)^{m_\sigma - 1} \frac{\mathcal{F}(p)\, r(p)}{\Lambda_\sigma(p)} \Big]_{p=\lambda_\sigma}. \tag{4.129}$$

Cette fonction est une matrice rectangulaire $n \times m$. Comme d'après le théorème de multiplication des transformées

$$\frac{1}{p} p\nu(p)\zeta(p) \div \int_0^t N(t-u) z(u) du, \qquad (4.130)$$

la loi du mouvement d'un système décrit par l'équation différentielle vectorielle (99) est

$$x(t) = M(t) + S(D) z(t) + \int_0^t N(t-u) z(u) du. \qquad (4.131)$$

Les éléments $x_j(t)$ du vecteur $x(t)$ seront

$$x_j(t) = M_j(t) + \sum_{\mu=1}^{m} S_{j\mu}(D) z_\mu(t) + \sum_{\mu=1}^{m} \int_0^t N_{j\mu}(t-u) z_\mu(u) du \qquad (4.132)$$

$$(j = 1, \ldots, n).$$

b) Matrice de poids. La matrice de transfert du système $\Phi(D)$ est déterminée plus haut par l'expression (106). La fonction $\Phi(D)$ est une matrice $n \times m$. D'après (106), l'élément $\Phi_{js}(p)$ de la matrice $\Phi(p)$ s'écrit

$$\Phi_{js}(p) = \frac{\sum_{k=1}^{n} F_{jk}(p) r_{ks}(p)}{\Delta(p)} = S_{js}(p) + \nu_{js}(p), \qquad (4.133)$$

où $S_{js}(p)$ est un polynôme en p et $\nu_{js}(p)$, une fraction régulière obtenue en séparant la partie entière de $\Phi_{js}(p)$.

Soient maintenant les éléments du vecteur $z(t)$ du second membre de l'équation vectorielle (99)

$$z_s(t) = \delta(t), \; z_q(t) \equiv 0, \; (q = 1, 2, \ldots, s-1, s+1, \ldots, m), \qquad (4.134)$$

où $\delta(t)$ est la fonction impulsion de Dirac. Sous la condition (134), le vecteur $r(D) z(t)$ est de la forme:

$$r(D) z(t) = \begin{bmatrix} r_{1s}(D) z_s(t) \\ r_{2s}(D) z_s(t) \\ \cdots\cdots \\ r_{ns}(D) z_s(t) \end{bmatrix} = \begin{bmatrix} r_{1s}(D) \delta(t) \\ r_{2s}(D) \delta(t) \\ \cdots\cdots \\ r_{ns}(D) \delta(t) \end{bmatrix}. \qquad (4.135)$$

D'une façon analogue à (22), la transformée de la matrice (135) s'écrit

$$\begin{bmatrix} r_{1s}(D)\,\delta(t) \\ r_{2s}(D)\,\delta(t) \\ \cdots\cdots \\ r_{ns}(D)\,\delta(t) \end{bmatrix} \risingdotseq \begin{bmatrix} pr_{1s}(p) \\ pr_{2s}(p) \\ \cdots \\ pr_{ns}(p) \end{bmatrix} = p \begin{bmatrix} r_{1s}(p) \\ r_{2s}(p) \\ \cdots \\ r_{ns}(p) \end{bmatrix}. \tag{4.136}$$

En introduisant les notations

$$\tilde{r}(p) = \begin{bmatrix} r_{1s}(p) \\ r_{2s}(p) \\ \cdots \\ r_{ns}(p) \end{bmatrix}, \tag{4.137}$$

on trouve que si les éléments du vecteur $z(t)$ sont de la forme (134),

$$r(D)\,z(t) \risingdotseq p\tilde{r}(p). \tag{4.138}$$

Nous sommes ainsi dans le cas où la transformation (111) devient

$$f(p)\,\xi(p) = pQ(p) + p\tilde{r}(p), \tag{4.139}$$

d'où l'on tire

$$\xi(p) = p\frac{F(p)\,Q(p)}{\Delta(p)} + p\frac{F(p)\,\tilde{r}(p)}{\Delta(p)}. \tag{4.140}$$

En tenant compte de (137), l'élément $\xi_j(p)$ du vecteur $\xi(p)$ s'écrit

$$\xi_j(p) = p\frac{\sum_{k=1}^{n} F_{jk}(p)\,Q_k(p)}{\Delta(p)} + p\frac{\sum_{k=1}^{n} F_{jk}(p)\,r_{ks}(p)}{\Delta(p)} \quad (j = 1, \ldots, n). \tag{4.141}$$

Conformément à (133), on peut récrire l'expression (141) sous la forme

$$\xi_j(p) = p\frac{\sum_{k=1}^{n} F_{jk}(p)\,Q_k(p)}{\Delta(p)} + pS_{js}(p) + p\nu_{js}(p) \quad (j = 1, \ldots, n). \tag{4.142}$$

D'après (110), dans les conditions initiales nulles

$$x_k(0) = 0,\ \dot{x}_k(0) = 0,\ \ldots,\ \overset{\nu_k - 1}{x_k}(0) = 0 \quad (k = 1, \ldots, n) \tag{4.143}$$

la matrice $Q(p)$ s'annule

$$Q(p) = 0, \tag{4.144}$$

et l'expression (4.142) devient

$$\xi_j(p) = pS_{js}(p) + p\nu_{js}(p) = p\Phi_{js}(p). \tag{4.145}$$

Etant donné que $S_{js}(p)$ est un polynôme en p,

$$pS_{js}(p) \doteq S_{js}(D)\,\delta(t). \tag{4.146}$$

Si l'on retient encore que d'après (128)

$$p\nu_{js}(p) \doteq N_{js}(t), \tag{4.147}$$

où $N_{js}(t)$ est un élément de la matrice $N(t)$ déterminée par l'expression (129), on trouve qu'à la transformée $p\Phi_{js}(p)$ correspond l'original

$$p\Phi_{js}(p) \doteq K_{js}(t), \tag{4.148}$$

avec

$$K_{js}(t) = S_{js}(D)\,\delta(t) + N_{js}(t). \tag{4.149}$$

La matrice $N(t)$ a été déterminée précédemment par l'expression (129). En ramenant l'expression de l'élément $N_{js}(t)$ de cette matrice à la forme réelle, on a

$$N_{js}(t) = \sum_{g=1}^{s'} \frac{e^{\varkappa_g t}}{(m_g-1)!}\left[\left(t+\frac{\partial}{\partial p}\right)^{m_g-1} \frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(p)\,r_{ks}(p)}{\Lambda_g(p)}\right]_{p=\varkappa_g} +$$

$$+2\sum_{h=1}^{s''} \frac{e^{\varepsilon_h t}}{(m_{s'+h}-1)!}\,[\operatorname{Re}\mathcal{L}\cos\omega_h t - \operatorname{Im}\mathcal{L}\sin\omega_h t], \tag{4.150}$$

où

$$\mathcal{L} = \left[\left(t+\frac{\partial}{\partial p}\right)^{m_{s'+h}-1} \frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(p)\,r_{ks}(p)}{\Lambda_{s'+h}(p)}\right]_{p=\varepsilon_h+i\omega_h}.$$

Ici s' est le nombre de racines réelles $\varkappa_g$ de l'équation $\Lambda(p) = 0$, où $\Lambda(p)$ est de la forme (123), qui ne coïncident pas; s'', le nombre de couples de racines complexes $\varepsilon_h \pm i\omega_h$ de cette équation qui ne coïncident pas non plus. Par m_g et $m_{s'+h}$ nous avons désigné la multiplicité de ces racines et par $\Lambda_g(p)$ et $\Lambda_{s'+h}(p)$, les fonctions

$$\Lambda_g(p) = \frac{\Lambda(p)}{(p-\varkappa_g)^{m_g}}, \quad \Lambda_{s'+h}(p) = \frac{\Lambda(p)}{(p-\varepsilon_h-i\omega_h)^{m_{s'+h}}}. \tag{4.151}$$

D'après (145) et (148)

$$x_j(t) = K_{js}(t) \quad (j = 1, \ldots, n). \tag{4.152}$$

Ainsi, la fonction $K_{js}(t)$ ($j = 1, \ldots, n$; s fixé), dans les conditions initiales nulles est une loi de variation dans le temps de la coordonnée x_j du système dans le cas où tous les éléments du signal

d'entrée vectoriel $z(t)$ sont identiquement nuls, sauf l'élément $z_s(t)$ qui vaut la fonction impulsion de Dirac $\delta(t)$.

La relation opérationnelle (148) entraîne

$$p\Phi(p) \doteqdot K(t), \tag{4.153}$$

où $K(t)$ est une matrice $m \times n$, dont l'élément $K_{js}(t)$ est défini par l'expression (149).

La matrice $K(t)$ définie par la relation opérationnelle (153) peut s'appeler *matrice de poids* du système multivariable décrit par l'équation différentielle vectorielle (99).

Conformément à (153), l'expression (149) ne définit la fonction $K(t)$ que pour $t \geqslant 0$. Comme pour $t < 0$ le système était en équilibre, nous devons adopter en vertu de (152)

$$K(t) = 0 \text{ pour } t < 0.$$

c) Matrice de transition. Considérons maintenant le cas où les éléments du vecteur $z(t)$ sont de la forme

$$z_s(t) = 1(t), \; z_q(t) \equiv 0 \quad (q = 1, 2, \ldots, s-1, s+1, \ldots, m). \tag{4.154}$$

Sous la condition (154), le produit des matrices $r(D)$ et $z(t)$ s'écrit

$$r(D)\,z(t) = \begin{bmatrix} r_{1s}(D)\,z_s(t) \\ r_{2s}(D)\,z_s(t) \\ \cdots\cdots \\ r_{ns}(D)\,z_s(t) \end{bmatrix} = \begin{bmatrix} r_{1s}(D)\,1(t) \\ r_{2s}(D)\,1(t) \\ \cdots\cdots \\ r_{ns}(D)\,1(t) \end{bmatrix}. \tag{4.155}$$

On a pour la transformée du vecteur (155)

$$\begin{bmatrix} r_{1s}(D)\,1(t) \\ r_{2s}(D)\,1(t) \\ \cdots\cdots \\ r_{ns}(D)\,1(t) \end{bmatrix} \doteqdot \begin{bmatrix} r_{1s}(p) \\ r_{2s}(p) \\ \cdots \\ r_{ns}(p) \end{bmatrix}, \tag{4.156}$$

c'est-à-dire

$$r(D)\,z(t) \doteqdot \tilde{r}(p), \tag{4.157}$$

où le vecteur $\tilde{r}(p)$ est défini par l'expression (137). Nous sommes ici dans le cas où la transformation (111) devient

$$f(p)\,\xi(p) = pQ(p) + \tilde{r}(p), \tag{4.158}$$

et

$$\xi(p) = p\,\frac{F(p)\,Q(p)}{\Delta(p)} + \frac{F(p)\,\tilde{r}(p)}{\Delta(p)}. \tag{4.159}$$

Pour l'élément $\xi_j(p)$ du vecteur $\xi(p)$ on a, compte tenu de (141),

$$\xi_j(p) = p\frac{\sum_{k=1}^{n} F_{jk}(p)\,Q_k(p)}{\Delta(p)} + \frac{\sum_{k=1}^{n} F_{jk}(p)\,r_{ks}'(p)}{\Delta(p)} \quad (j=1, \ldots, n). \quad (4.160)$$

D'après (133), l'expression (160) peut s'écrire ainsi:

$$\xi_j(p) = p\frac{\sum_{k=1}^{n} F_{jk}(p)\,Q_k(p)}{\Delta(p)} + S_{js}(p) + \nu_{js}(p). \quad (4.161)$$

Les conditions initiales (143) étant nulles, la matrice $Q(p)$ s'annule et l'expression (161) devient

$$\xi_j(p) = S_{js}(p) + \nu_{js}(p) = \Phi_{js}(p) \quad (j = 1, 2, \ldots, n). \quad (4.162)$$

Puisque $S_{js}(p)$ est un polynôme en p, on a

$$S_{js}(p) \div S_{js}(D)\,1(t). \quad (4.163)$$

Désignons par $A(t)$ la matrice dont la transformée est la matrice $\nu(p)$:

$$\nu(p) \div A(t). \quad (4.164)$$

D'une façon analogue à (88), il vient

$$A(t) = \frac{\mathcal{F}(0)\,r(0)}{\Lambda(0)} + \sum_{\sigma=1}^{s} \frac{e^{\lambda_\sigma t}}{(m_\sigma - 1)!}\left[\left(t + \frac{\partial}{\partial p}\right)^{m_\sigma - 1} \frac{\mathcal{F}(p)\,r(p)}{p\Lambda_\sigma(p)}\right]_{p=\lambda_\sigma}. \quad (4.165)$$

(164) entraîne que

$$\nu_{js}(p) \div A_{js}(t), \quad (4.166)$$

où la fonction $A_{js}(t)$ est un élément de la matrice $A(t)$.

D'une façon analogue à (150), l'expression de l'élément $A_{js}(t)$ de la matrice $A(t)$ ramenée à une forme réelle est

$$A_{js}(t) = \frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(0)\,r_{ks}(0)}{\Lambda(0)} +$$

$$+ \sum_{g=1}^{s'} \frac{e^{\varkappa_g t}}{(m_g - 1)!}\left[\left(t + \frac{\partial}{\partial p}\right)^{m_g - 1} \frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(p)\,r_{ks}(p)}{p\Lambda_g(p)}\right]_{p=\varkappa_g} +$$

$$+ 2\sum_{h=1}^{s''} \frac{e^{\varepsilon_h t}}{(m_{s'+h} - 1)!} \times$$

$$\times\left\{\operatorname{Re}\left[\left(t+\frac{\partial}{\partial p}\right)^{m_{s'+h}-1}\frac{\sum_{k=1}^{n}\mathcal{F}_{jk}(p)\,r_{ks}(p)}{p\Lambda_{s'+h}(p)}\right]_{p=\varepsilon_h+i\omega_h}\cos\omega_h t-\right.$$

$$\left.-\operatorname{Im}\left[\left(t+\frac{\partial}{\partial p}\right)^{m_{s'+h}-1}\frac{\sum_{k=1}^{n}\mathcal{F}_{jk}(p)\,r_{ks}(p)}{p\Lambda_{s'+h}(p)}\right]_{p=\varepsilon_h+i\omega_h}\sin\omega_h t\right\}. \quad (4.167)$$

Ainsi, à la transformée $\Phi_{js}(p)$ correspond l'original

$$\Phi_{js}(p) \div U_{js}(t), \quad (4.168)$$

où

$$U_{js}(t) = S_{js}(D)\,1(t) + A_{js}(t). \quad (4.169)$$

En vertu de (162) et (168)

$$x_j(t) = U_{js}(t) \quad (j = 1, \ldots, n). \quad (4.170)$$

Ainsi, la fonction $U_{js}(t)$ ($j = 1, \ldots, n$; s fixé) est une loi qui régit la variation dans le temps de la coordonnée x_j du système dans les conditions initiales nulles pour le cas où tous les éléments du signal d'entrée vectoriel $z(t)$ sont identiquement nuls, sauf $z_s(t)$ qui est égal à l'échelon unitaire $1(t)$.

La relation opérationnelle (168) entraîne

$$\Phi(p) \div U(t), \quad (4.171)$$

où $U(t)$ est une matrice $m \times n$ dont l'élément $U_{js}(t)$ est déterminé par l'expression (169).

La matrice $U(t)$, définie par la relation opérationnelle (171), peut être nommée *matrice de transition* d'un système de commande multivariable décrit par l'équation différentielle vectorielle (99).

Notons que puisque tous les éléments de la matrice $\nu(p)$ sont des fractions régulières, $\lim_{p\to\infty}\nu(p)=0$ et, par conséquent, $A(0)=0$. Donc

$$\dot{A}(t) \div p\,[\nu(p) - A(0)] = p\nu(p). \quad (4.172)$$

En tenant encore compte du fait que

$$\frac{d}{dt}[S(D)\,1(t)] = S(D)\,\delta(t) \div pS(p), \quad (4.173)$$

on a

$$\dot{U}(t) \div pS(p) + p\nu(p) = p\Phi(p), \quad (4.174)$$

ce qui coïncide avec la transformée (153) de la fonction $K(t)$. On en déduit la relation suivante entre la matrice de poids $K(t)$ et la matrice de transition $U(t)$:

$$\dot{U}(t) = K(t). \quad (4.175)$$

§ 5. Processus transitoires et stationnaires dans les systèmes commandés en boucle fermée

1. Calcul de la fonction de poids d'après le lieu de transfert d'un système en boucle fermée. Au § 4 nous avons introduit à l'aide des relations

$$p\Phi_{js}(p) = p\,\frac{\sum\limits_{k=1}^{n} F_{jk}(p)\,r_{ks}(p)}{\Delta(p)} \div N_{js}(t), \qquad (5.1)$$

$$\Phi_{js}(p) = \frac{\sum\limits_{k=1}^{n} F_{jk}(p)\,r_{ks}(p)}{\Delta(p)} \div A_{js}(t) \qquad (5.2)$$

la fonction de poids (4.148), (4.133) (qui s'appelle également réponse impulsionnelle) et la réponse unitaire (4.168). Nous supposerons ici que la fonction $\Phi_{js}(p)$ est une fraction régulière. En nous bornant aux systèmes *stables* nous devons poser que tous les zéros du polynôme $\Delta(p)$ se situent dans le demi-plan p gauche. C'est pourquoi tous les pôles de la fonction $\Phi_{js}(p)$ reposent dans ce demi-plan gauche.

Si l'on connaît la fonction $\Phi_{js}(p)$, alors la fonction de poids $N_{js}(t)$ et la réponse unitaire $A_{js}(t)$ peuvent s'obtenir à l'aide de la formule de Riemann-Mellin. La relation (1) conduit à

$$N_{js}(t) = \frac{1}{2\pi i}\int\limits_{c-i\infty}^{c+i\infty} \frac{p\Phi_{js}(p)}{p}\, e^{pt}\, dp \quad (t \geqslant 0)$$

ou

$$N_{js}(t) = \frac{1}{2\pi i}\int\limits_{c-i\infty}^{c+i\infty} \Phi_{js}(p)\, e^{pt}\, dp \quad (t \geqslant 0). \qquad (5.3)$$

Dans l'expression (3) l'intégrale est calculée le long d'une droite parallèle à l'axe imaginaire (menée à la distance c de cet axe), la grandeur c étant choisie de façon que tous les pôles de la fonction $\Phi_{js}(p)$ se trouvent à gauche de la droite envisagée. Tous les pôles de la fonction $\Phi_{js}(p)$ se trouvant dans le demi-plan gauche, on peut poser $c = 0$.

Passons maintenant du plan de la variable complexe p dans le plan q où

$$p = iq. \qquad (5.4)$$

La formule (4) amène

$$q = -ip.$$

Ainsi, les relations

$$p = c - i\infty, \quad q = -i(c - i\infty) = -ic - \infty,$$
$$p = c + i\infty, \quad q = -i(c + i\infty) = -ic + \infty,$$

sont vraies et l'expression (3) s'écrit

$$N_{js}(t) = \frac{1}{2\pi} \int_{-ic-\infty}^{-ic+\infty} \Phi_{js}(iq)\, e^{iqt}\, dq. \tag{5.5}$$

Puisque

$$p = c + i\omega, \tag{5.6}$$

on a

$$q = -i(c + i\omega) = -ic + \omega. \tag{5.7}$$

Avec $c = 0$, l'expression (7) devient

$$q = \omega.$$

Par conséquent, en vertu de l'hypothèse avancée plus haut suivant laquelle le système concerné est stable, on peut poser $c = 0$ et ramener l'expression (5) à la forme

$$N_{js}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{js}(i\omega)\, e^{i\omega t}\, d\omega. \tag{5.8}$$

L'expression (5.8) est une transformation inverse de Fourier. Elle définit la fonction $N_{js}(t)$ dans l'intervalle $0 < t < \infty$, si seulement l'intégrale $\int_{-\infty}^{\infty} N_{js}(t)\, dt$ est absolument convergente. Pour un système stable cette condition est vérifiée.

L'expression (8) peut s'écrire

$$N_{js}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\operatorname{Re} \Phi_{js}(i\omega) \cos \omega t - \operatorname{Im} \Phi_{js}(i\omega) \sin \omega t]\, d\omega +$$
$$+ \frac{i}{2\pi} \int_{-\infty}^{\infty} [\operatorname{Re} \Phi_{js}(i\omega) \sin \omega t + \operatorname{Im} \Phi_{js}(i\omega) \cos \omega t]\, d\omega. \tag{5.9}$$

Les fonctions $\operatorname{Re} \Phi_{js}(i\omega)$ et $\cos \omega t$ sont des fonctions paires de ω. Les fonctions $\operatorname{Im} \Phi_{js}(i\omega)$ et $\sin \omega t$ sont des fonctions impaires de ω. Donc la fonction sous le signe somme de la deuxième intégrale de l'expression (9) est une fonction impaire de ω, et comme l'intégration s'effectue entre $-\infty$ et $+\infty$, la deuxième intégrale de (9) s'annule. La fonction sous le signe somme de la première intégrale de (9) étant

une fonction paire de ω, cette intégrale peut être remplacée par le double de cette intégrale prise dans les limites de 0 à ∞. L'expression (9) s'écrit alors

$$N_{js}(t)=\frac{1}{\pi}\int_0^\infty \operatorname{Re}\Phi_{js}(i\omega)\cos\omega t\,d\omega-\frac{1}{\pi}\int_0^\infty \operatorname{Im}\Phi_{js}(i\omega)\sin\omega t\,d\omega. \quad (5.10)$$

La formule (10) permet ainsi de définir la fonction de poids $N_{js}(t)$ du système d'après le lieu de transfert $\Phi_{js}(i\omega)$ de ce système.

Soit, maintenant, $t=-\tau$. L'expression (8) devient alors

$$N_{js}(-\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_{js}(i\omega)\,e^{-i\omega\tau}\,d\omega. \quad (5.11)$$

En revenant dans le plan p, on a en vertu de (6)

$$N_{js}(-\tau)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Phi_{js}(p)\,e^{-p\tau}\,dp. \quad (5.12)$$

Fig. 5.1

Puisque d'après l'hypothèse que nous venons de faire $\Phi_{js}(p)$ est une fraction régulière, sur le demi-cercle Γ de rayon R (fig. 5.1) la fonction $\Phi_{js}(p)$, où $p=Re^{i\varphi}$, avec $R\to\infty$, tend uniformément vers zéro par rapport à φ, où $-\frac{\pi}{2}<\varphi<\frac{\pi}{2}$. C'est pourquoi ce demi-cercle vérifie le lemme de Jordan

$$\lim_{R\to\infty}\frac{1}{2\pi i}\int_{\Gamma}\Phi_{js}(p)\,e^{-p\tau}\,dp=0. \quad (5.13)$$

Le demi-cercle Γ d'intégration dans l'expression (13) repose dans le demi-plan p droit du fait que pour $p=Re^{i\varphi}$ c'est seulement dans le demi-cercle p droit que la fonction $|e^{-p\tau}|=e^{-(\operatorname{Re}p)\tau}$ tend vers zéro avec $R\to\infty$. La relation (13) permet d'écrire l'expression (12) sous la forme

$$N_{js}(-\tau)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Phi_{js}(p)\,e^{-p\tau}\,dp+\lim_{R\to\infty}\frac{1}{2\pi i}\int_{\Gamma}\Phi_{js}(p)\,e^{-p\tau}\,dp=$$
$$=\frac{1}{2\pi i}\oint_C\Phi_{js}(p)\,e^{-p\tau}\,dp, \quad (5.14)$$

où le contour fermé C est composé de l'axe imaginaire et d'un demi-cercle de rayon infiniment grand. Ce contour renferme tout le demi-plan droit.

L'intégrale du second membre de (14) est égale à la somme des résidus de la fonction $\Phi_{js}(p)\, e^{-p\tau}$ suivant les points singuliers de la fonction $\Phi_{js}(p)$ situés dans le demi-plan p droit. Mais comme dans le demi-plan droit la fonction $\Phi_{js}(p)$ ne possède aucun point singulier, cette intégrale est nulle. Ainsi on obtient que

$$N_{js}(-\tau) = 0. \tag{5.15}$$

Puisque d'après (10)

$$N_{js}(-\tau) = \frac{1}{\pi}\int_0^\infty \operatorname{Re}\Phi_{js}(i\omega)\cos\omega\tau\, d\omega + \frac{1}{\pi}\int_0^\infty \operatorname{Im}\Phi_{js}(i\omega)\sin\omega\tau\, d\omega, \tag{5.16}$$

conformément à (15) on a

$$\frac{1}{\pi}\int_0^\infty \operatorname{Re}\Phi_{js}(i\omega)\cos\omega\tau\, d\omega = -\frac{1}{\pi}\int_0^\infty \operatorname{Im}\Phi_{js}(i\omega)\sin\omega\tau\, d\omega. \tag{5.17}$$

Enfin (17) et (10) conduisent à l'expression définitive de la fonction de poids

$$N_{js}(t) = \frac{2}{\pi}\int_0^\infty \operatorname{Re}\Phi_{js}(i\omega)\cos\omega t\, d\omega =$$

$$= -\frac{2}{\pi}\int_0^\infty \operatorname{Im}\Phi_{js}(i\omega)\sin\omega t\, d\omega \quad \text{pour } t \geqslant 0. \tag{5.18}$$

2. Calcul de la réponse unitaire d'après le lieu de transfert d'un système en boucle fermée. Passons maintenant à la réponse unitaire $A_{js}(t)$. Puisque d'après (2)

$$\Phi_{js}(p) = \frac{\sum\limits_{k=1}^{n} F_{jk}(p)\, r_{ks}(p)}{\Delta(p)} \div A_{js}(t), \tag{5.19}$$

il vient

$$A_{js}(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{\Phi_{js}(p)}{p} e^{pt}\, dp \quad (t \geqslant 0). \tag{5.20}$$

Il est impossible d'appliquer directement à l'expression (20) les résultats obtenus ci-dessus du fait que la fonction $\frac{\Phi_{js}(p)}{p} = \frac{\sum\limits_{k=1}^{n} F_{jk}(p) r_{ks}(p)}{p\Delta(p)}$ possède en origine un pôle de premier ordre.

C'est pourquoi nous allons transformer cette fonction, et à cet effet, la développer en une somme de fractions élémentaires. En tenant compte des relations (4.122), on a

$$\frac{\Phi_{js}(p)}{p}=\frac{\sum_{k=1}^{n} F_{jk}(p)\,r_{ks}(p)}{p\Delta(p)}=\frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(p)\,r_{ks}(p)}{p\Lambda(p)}=$$

$$=\frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(0)\,r_{ks}(0)}{\Lambda(0)}\,\frac{1}{p}+\sum_{\sigma=1}^{s}\sum_{\rho=0}^{m_\sigma-1}\frac{1}{(m_\sigma-\rho-1)!}\times$$

$$\times\left[\frac{\partial^{m_\sigma-\rho-1}}{\partial p^{m_\sigma-\rho-1}}\,\frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(p)\,r_{ks}(p)}{p\Lambda_\sigma(p)}\right]_{p=\lambda_\sigma}\frac{1}{(p-\lambda_\sigma)^{\rho+1}}. \qquad (5.21)$$

Introduisons la notation $\Psi_{js}(p)$ pour le deuxième terme du second membre de (21). Ainsi,

$$\frac{\Phi_{js}(p)}{p}=\frac{\sum_{k=1}^{n} \mathcal{F}_{jk}(0)\,r_{ks}(0)}{\Lambda(0)}\,\frac{1}{p}+\Psi_{js}(p)$$

ou

$$\frac{\Phi_{js}(p)}{p}=\Phi_{js}(0)\,\frac{1}{p}+\Psi_{js}(p). \qquad (5.22)$$

$\Psi_{js}(p)$ est une fonction rationnelle dont tous les pôles se situent à gauche de l'axe imaginaire dans le plan de la variable complexe p.

Puisque d'après (22)

$$\Psi_{js}(p)=\frac{\Phi_{js}(p)-\Phi_{js}(0)}{p}, \qquad (5.23)$$

on écrira dans ce qui suit

$$\frac{\Phi_{js}(p)}{p}=\frac{\Phi_{js}(0)}{p}+\frac{\Phi_{js}(p)-\Phi_{js}(0)}{p}, \qquad (5.24)$$

tous les pôles du deuxième terme de l'expression (24) reposant dans le demi-plan p gauche.

Maintenant l'expression (20) peut s'écrire

$$A_{js}(t)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\Phi_{js}(0)}{p}\,e^{pt}\,dp+\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\Phi_{js}(p)-\Phi_{js}(0)}{p}\,e^{pt}\,dp. \qquad (5.25)$$

Comme $\Phi_{js}(0)$ est une constante, $\Phi_{js}(0) \div\!\!\rightarrow \Phi_{js}(0)$ (les constantes se confondent avec leurs transformées) et

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\Phi_{js}(0)}{p}\,e^{pt}\,dp=\Phi_{js}(0). \qquad (5.26)$$

L'expression (25) devient

$$A_{js}(t)=\Phi_{js}(0)+\frac{1}{2\pi i}\int\limits_{c-i\infty}^{c+i\infty}\frac{\Phi_{js}(p)-\Phi_{js}(0)}{p}e^{pt}\,dp. \qquad (5.27)$$

Le deuxième terme du second membre de (27) peut être ramené à une forme analogue à celle de l'expression (18). Calculons au préalable les parties réelle et imaginaire de la fonction

$$\frac{\Phi_{js}(i\omega)-\Phi_{js}(0)}{i\omega}.$$

On a

$$\frac{\Phi_{js}(i\omega)-\Phi_{js}(0)}{i\omega}=\frac{\operatorname{Re}\Phi_{js}(i\omega)-\Phi_{js}(0)}{i\omega}+i\,\frac{\operatorname{Im}\Phi_{js}(i\omega)}{i\omega}.$$

On en tire

$$\left.\begin{aligned}\operatorname{Re}\left[\frac{\Phi_{js}(i\omega)-\Phi_{js}(0)}{i\omega}\right]&=\frac{\operatorname{Im}\Phi_{js}(i\omega)}{\omega},\\ \operatorname{Im}\left[\frac{\Phi_{js}(i\omega)-\Phi_{js}(0)}{i\omega}\right]&=\frac{\Phi_{js}(0)-\operatorname{Re}\Phi_{js}(i\omega)}{\omega}.\end{aligned}\right\} \qquad (5.28)$$

Donc

$$\frac{1}{2\pi i}\int\limits_{c-i\infty}^{c+i\infty}\frac{\Phi_{js}(p)-\Phi_{js}(0)}{p}e^{pt}\,dp=\frac{2}{\pi}\int\limits_{0}^{\infty}\frac{\operatorname{Im}\Phi_{js}(i\omega)}{\omega}\cos\omega t\,d\omega=$$

$$=\frac{2}{\pi}\int\limits_{0}^{\infty}\left[\frac{\operatorname{Re}\Phi_{js}(i\omega)}{\omega}-\frac{\Phi_{js}(0)}{\omega}\right]\sin\omega t\,d\omega. \qquad (5.29)$$

En portant cette expression dans (27), on obtient

$$A_{js}(t)=\Phi_{js}(0)+\frac{2}{\pi}\int\limits_{0}^{\infty}\frac{\operatorname{Im}\Phi_{js}(i\omega)}{\omega}\cos\omega t\,d\omega=$$

$$=\Phi_{js}(0)-\frac{2}{\pi}\int\limits_{0}^{\infty}\frac{\Phi_{js}(0)}{\omega}\sin\omega t\,d\omega+\frac{2}{\pi}\int\limits_{0}^{\infty}\frac{\operatorname{Re}\Phi_{js}(i\omega)}{\omega}\sin\omega t\,d\omega. \qquad (5.30)$$

Etant donné que

$$\int\limits_{0}^{\infty}\frac{\sin\omega t}{\omega}\,d\omega=\frac{\pi}{2}, \qquad (5.31)$$

l'expression (30) se met sous la forme

$$A_{js}(t)=\frac{2}{\pi}\int_0^\infty \frac{\operatorname{Re}\Phi_{js}(i\omega)}{\omega}\sin\omega t\,d\omega=$$

$$=\Phi_{js}(0)+\frac{2}{\pi}\int_0^\infty \frac{\operatorname{Im}\Phi_{js}(i\omega)}{\omega}\cos\omega t\,d\omega. \quad (5.32)$$

L'expression (32) a lieu pour $t\geqslant 0$.

3. Systèmes à déphasage minimal. Pour un système commandé monovariable la relation (18) s'écrit

$$N(t)=\frac{2}{\pi}\int_0^\infty \operatorname{Re}\Phi(i\omega)\cos\omega t\,d\omega=-\frac{2}{\pi}\int_0^\infty \operatorname{Im}\Phi(i\omega)\sin\omega t\,d\omega \quad (t\geqslant 0). \quad (5.33)$$

En adoptant les notations

$$\operatorname{Re}\Phi(i\omega)=P(\omega),\quad \operatorname{Im}\Phi(i\omega)=Q(\omega), \quad (5.34)$$

on peut récrire la relation (33) de la façon suivante:

$$N(t)=\frac{2}{\pi}\int_0^\infty P(\omega)\cos\omega t\,d\omega=-\frac{2}{\pi}\int_0^\infty Q(\omega)\sin\omega t\,d\omega. \quad (5.35)$$

D'une façon analogue, pour un système monovariable la relation (32) devient

$$A(t)=\frac{2}{\pi}\int_0^\infty \frac{P(\omega)}{\omega}\sin\omega t\,d\omega=P(0)+\frac{2}{\pi}\int_0^\infty \frac{Q(\omega)}{\omega}\cos\omega t\,d\omega. \quad (5.36)$$

La relation (36) rend compte du fait que

$$\Phi(0)=[\Phi(p)]_{p=0}=P(0).$$

En déduisant les relations (18) et (32), nous avons admis que la fonction $\Phi(p)$, qui est une fraction rationnelle de p:

$$\Phi(p)=\frac{l(p)}{f(p)}, \quad (5.37)$$

est une fraction régulière et que le système concerné est *asymptotiquement stable*. Ceci signifie que tous les zéros du polynôme $f(p)$ ou, ce qui revient au même, tous les pôles de la fonction $\Phi(p)$ se situent dans le plan de la variable complexe p strictement à gauche de l'axe imaginaire.

Remarquons sans démonstration ([82], [41], p. 246) que pour les systèmes asymptotiquement stables (c'est-à-dire pour la fonction $\Phi(p)$ dont tous les pôles reposent dans le demi-plan p gauche), les fonctions $P(\omega)$ et $Q(\omega)$ se définissent mutuellement étant liées par des formules suivantes qui représentent la *transformation de Hilbert*:

$$\left.\begin{aligned} P(\omega) &= -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{Q(u)}{u-\omega}\,du, \\ Q(\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{P(u)}{u-\omega}\,du. \end{aligned}\right\} \tag{5.38}$$

En désignant par $R(\omega)$ le module de la fonction $\Phi(i\omega)$, et par $\psi(\omega)$ son argument, on a

$$\Phi(i\omega) = R(\omega)\,e^{i\psi(\omega)}, \tag{5.39}$$

ce qui conduit à

$$\ln\Phi(i\omega) = \ln R(\omega) + i\psi(\omega). \tag{5.40}$$

Les fonctions $\ln R(\omega)$ et $\psi(\omega)$ sont associées à la fonction $\ln\Phi(i\omega)$ de la même façon que $P(\omega)$ et $Q(\omega)$ le sont à la fonction $\Phi(i\omega)$. C'est pourquoi on peut dire d'une façon analogue à (38) que *si tous les pôles de la fonction* $\ln\Phi(p)$ *reposent strictement à gauche de l'axe imaginaire dans le plan de la variable complexe* p, les fonctions $\ln R(\omega)$ et $\psi(\omega)$ vérifient les relations

$$\left.\begin{aligned} \ln R(\omega) &= -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\psi(u)}{u-\omega}\,du, \\ \psi(\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\ln R(u)}{u-\omega}\,du. \end{aligned}\right\} \tag{5.41}$$

Etant donné que d'après (37)

$$\ln\Phi(p) = \ln\frac{l(p)}{f(p)} = \ln l(p) - \ln f(p), \tag{5.42}$$

parmi les points singuliers de la fonction $\ln\Phi(p)$ il y a non seulement les zéros de la fonction $f(p)$ mais aussi ceux de la fonction $l(p)$. (En vertu de (42), si $l(p)\to 0$, on a $\ln\Phi(p)\to -\infty$.)

Ainsi, les formules (41) qui définissent $R(\omega)$ à l'aide de $\psi(\omega)$ et inversement ne sont vérifiées que si *non seulement tous les pôles de la fonction*

$$\Phi(p) = \frac{l(p)}{f(p)}$$

mais encore tous ses zéros se situent strictement à gauche de l'axe imaginaire dans le plan de la variable complexe p. Les systèmes qui respectent cette condition s'appellent *systèmes à déphasage minimal.*

Les systèmes à déphasage minimal se distinguent par le fait que de tous les systèmes possibles de même réponse en amplitude $R(\omega)$ ils ont le plus petit déphasage quelle que soit la valeur de la fréquence ω. Par *déphasage* on entend la valeur *négative* de l'argument de $\Phi(i\omega)$, c'est-à-dire $-\psi(\omega)$.

Ainsi, la fonction

$$\Phi_1(p)=\frac{(1+p\tau_1)(1+p\tau_2)}{(1+pT_1)(1+pT_2)} \tag{5.43}$$

est la fonction de transfert d'un système à déphasage minimal, alors que

$$\Phi_2(p)=\frac{(1-p\tau_1)(1+p\tau_2)}{(1+pT_1)(1+pT_2)} \tag{5.44}$$

est la fonction de transfert d'un système à déphasage non minimal.

La fonction $\Phi_2(p)$ peut s'écrire

$$\Phi_2(p)=\Phi_1(p)\frac{1-p\tau_1}{1+p\tau_1}. \tag{5.45}$$

D'où

$$\Phi_2(i\omega)=\Phi_1(i\omega)\frac{1-i\omega\tau_1}{1+i\omega\tau_1}. \tag{5.46}$$

Puisque

$$\frac{1-i\omega\tau_1}{1+i\omega\tau_1}=e^{-2i\varphi},\quad \operatorname{tg}\varphi=\omega\tau_1, \tag{5.47}$$

il vient

$$\Phi_2(i\omega) = \Phi_1(i\omega)\, e^{-2i\varphi}. \tag{5.48}$$

Introduisons les notations

$$\mu_1(\omega) = -\arg \Phi_1(i\omega). \tag{5.49}$$

Nous avons dit plus haut que la fonction $\mu_1(\omega)$ s'appelle *déphasage.* Voici le sens de cette notion. Le signal de sortie d'un système à fonction de transfert $\Phi(D)$ est

$$x = \Phi(D)\, z(t).$$

Pour $z(t) = ae^{i\nu t}$ le processus stationnaire est traduit par l'expression

$$x(t)=\Phi(i\nu)\, ae^{i\nu t}=R(\nu)\, e^{-i\mu_1(\nu)}ae^{i\nu t}=R(\nu)\, ae^{i[\nu t-\mu_1(\nu)]}.$$

Comme il résulte de (48)

$$\arg \Phi_2(i\omega) = \arg \Phi_1(i\omega) - 2\varphi$$

ou

$$-\mu_2(\omega) = -\mu_1(\omega) - 2\varphi.$$

Par suite,

$$\mu_2(\omega) = \mu_1(\omega) + 2\varphi. \qquad (5.50)$$

De cette façon pour la même forme de la fonction $R(\omega)$ le décalage de la phase $\mu_1(\omega)$ du système à déphasage minimal est plus petit que celui de tout système à déphasage non minimal.

Les *systèmes à déphasage non minimal* avec zéros $p = \pm ik$ sur l'axe imaginaire s'emploient, par exemple, pour égaler à zéro $|\Phi(i\omega)|_{\omega=k}$, lorsqu'il faut que le système ne laisse pas passer des signaux de fréquence $\omega = k$. On associe également aux systèmes à déphasage non minimal les organes sans erreur de position (organes astatiques), c'est-à-dire les organes dont la fonction de transfert possède un pôle en origine.

Pour certaines formes de lieux de transfert d'un système en boucle ouverte, lorsqu'on peut rendre impossible l'encerclement par ce système du point $(-1, i0)$ par rotation du lieu de transfert dans le sens des aiguilles d'une montre, on branche en série dans le système en boucle ouverte (pour stabiliser le système en boucle fermée) un organe à déphasage non minimal.

Dans la synthèse des circuits électriques, les organes à déphasage non minimal s'emploient comme correcteurs de phase. Les systèmes à déphasage non minimal trouvent également d'autres applications.

4. Processus stationnaires dans les systèmes en boucle fermée. Constantes d'erreur. L'équation différentielle qui décrit les processus ayant lieu dans un système en boucle fermée monovariable a été donnée au § 4 sous la forme suivante:

$$f(D)\,x = l(D)\,z(t), \qquad (5.51)$$

où $D = \frac{d}{dt}$ et les fonctions $f(D)$ et $l(D)$ sont des polynômes de l'opérateur de différentiation $D = \frac{d}{dt}$ à coefficients constants

$$f(D) = a_0 D^n + a_1 D^{n-1} + \ldots + a_{n-1} D + a_n, \qquad (5.52)$$
$$l(D) = l_0 D^m + l_1 D^{m-1} + \ldots + l_{m-1} D + l_m \quad (m < n).$$

La fonction de transfert d'un système *en boucle fermée* est

$$\Phi(D) = \frac{l(D)}{f(D)}. \qquad (5.53)$$

Comme nous l'avons indiqué au § 1, le processus stationnaire est la loi du mouvement d'un système asymptotiquement stable dans le cas où le signal d'entrée $z(t)$ est amené à l'instant $t_0 = -\infty$. En supposant que le système (51) soit stable, le processus stationnaire s'écrit en vertu des formules (1.22) et (4.19)

$$x(t) = \int_{-\infty}^{t} N(t-\tau)\,z(\tau)\,d\tau = \int_{0}^{\infty} N(\xi)\,z(t-\xi)\,d\xi, \qquad (5.54)$$

où la fonction de poids du système $N(t)$ est déterminée par la relation opérationnelle

$$p\Phi(p) \doteq N(t). \tag{5.55}$$

Dans le cas où le signal d'entrée $z(t)$ est un *polynôme* en t de degré q, la fonction $z(t-\xi)$ peut être mise sous la forme:

$$z(t-\xi)=\sum_{r=0}^{q}(-1)^r\frac{z^{(r)}(t)}{r!}\xi^r. \tag{5.56}$$

Dans ce cas l'expression (54) se ramène à la forme:

$$x(t)=\sum_{r=0}^{q}(-1)^r\frac{z^{(r)}(t)}{r!}\int_0^\infty \xi^r N(\xi)\,d\xi. \tag{5.57}$$

Les intégrales impropres de (57) sont les moments de la fonction de poids $N(t)$:

$$\mu_r=\int_0^\infty \xi^r N(\xi)\,d\xi \quad (r=0, 1, 2, \ldots). \tag{5.58}$$

Comme on le voit de l'expression (4.15) de la fonction $N(t)$, les intégrales (58) sont convergentes. Compte tenu du fait que conformément à la relation opérationnelle (55)

$$\int_0^\infty N(\xi)\,e^{-p\xi}\,d\xi=\Phi(p), \tag{5.59}$$

on trouve que

$$(-1)^r\int_0^\infty \xi^r N(\xi)\,e^{-p\xi}\,d\xi=\Phi^{(r)}(p). \tag{5.60}$$

Les relations (58), (59) et (60) entraînent

$$\mu_0=\int_0^\infty N(\xi)\,d\xi=\Phi(0), \tag{5.61}$$

$$\mu_r=\int_0^\infty \xi^r N(\xi)\,d\xi=(-1)^r\Phi^{(r)}(0) \quad (r=1, 2, \ldots). \tag{5.62}$$

En vertu de (61) et (62), l'expression (57) devient

$$x(t)=\sum_{r=0}^{q}\frac{\Phi^{(r)}(0)}{r!}z^{(r)}(t). \tag{5.63}$$

Supposons maintenant que le système décrit par l'équation (51) consiste à reproduire la fonction $\sigma(D)\,z(t)$ où $\sigma(D)$ est un polynôme de l'opérateur $D = d/dt$ à coefficients constants

$$\sigma(D)\,z(t) = \sigma_l z^{(l)}(t) + \sigma_{l-1} z^{(l-1)}(t) + \ldots + \sigma_1 z'(t) + \sigma_0 z(t). \quad (5.64)$$

Ainsi, l'erreur du système est

$$\varepsilon(t) = x(t) - \sigma(D)\,z(t). \quad (5.65)$$

Avec les relations (63), (64) et (65) on a

$$\varepsilon(t) = \sum_{r=0}^{l} \left[\frac{1}{r!}\Phi^{(r)}(0) - \sigma_r\right] z^{(r)}(t) + \sum_{r=l+1}^{q} \frac{1}{r!}\Phi^{(r)}(0)\, z^{(r)}(t). \quad (5.66)$$

Les grandeurs

$$c_r = \frac{1}{r!}\Phi^{(r)}(0) - \sigma_r \quad (r = 0, 1, 2, \ldots) \quad (5.67)$$

sont dites *constantes d'erreur*. Si le signal d'entrée $z(t)$ est un polynôme en t de degré égal ou inférieur à k, on a pour l'erreur $\varepsilon(t)$ du système

$$\varepsilon(t) = \sum_{r=0}^{k} c_r z^{(r)}(t). \quad (5.68)$$

Le coefficient

$$c_0 = \Phi(0) - \sigma_0 \quad (5.69)$$

s'appelle *constante d'erreur statique ou erreur de position*. Lorsque $c_0 = 0$, le système commandé est dit *astatique* ou *sans erreur de position*.

Le coefficient

$$c_1 = \Phi'(0) - \sigma_1 \quad (5.70)$$

se nomme *constante d'erreur de vitesse*. Le coefficient

$$c_2 = \frac{1}{2}\Phi''(0) - \sigma_2, \quad (5.71)$$

est dit *constante d'erreur d'accélération*. Par *système astatique d'ordre k* on entend un système dont

$$c_0 = 0,\ c_1 = 0,\ \ldots,\ c_{k-1} = 0,\ c_k \neq 0.$$

5. **Processus stationnaires dans un système asservi.** Examinons maintenant les constantes d'erreur d'un système asservi. Un système asservi doit reproduire le signal d'entrée $z(t)$. Soit $z(t)$ un certain polynôme en t. L'erreur d'un système asservi est déterminée par l'expression

$$\varepsilon(t) = x(t) - z(t). \quad (5.72)$$

D'après (64), on a pour un système asservi

$$\sigma_0 = 1,\ \sigma_1 = \sigma_2 = \ldots = 0. \quad (5.73)$$

Conformément à (67) et à (73), les constantes d'erreur d'un système asservi sont

$$c_0 = \Phi(0) - 1,$$
$$c_r = \frac{1}{r!}\Phi^{(r)}(0) \quad (r = 1, 2, \ldots). \tag{5.74}$$

Dans un système asservi *astatique* la constante c_0 doit être nulle, c'est-à-dire il faut que soit vérifiée la relation

$$\Phi(0) = 1. \tag{5.75}$$

On peut démontrer que pour qu'un système asservi en boucle fermée soit un système astatique de degré k (c'est-à-dire que ses constantes d'erreur $c_0, c_1, c_2, \ldots, c_{k-1}$ soient nulles, alors que $c_k \neq 0$), il faut et il suffit que la fonction de transfert $G(D)$ *en boucle ouverte* soit de la forme

$$G(D) = \frac{1}{D^k} G_1(D), \quad G_1(0) \neq 0. \tag{5.76}$$

Voici la démonstration de cette proposition. En vertu de (1.66), la fonction de transfert en boucle fermée s'écrit

$$\Phi(D) = \frac{G(D)}{1 + G(D)}. \tag{5.77}$$

D'après (76) et (77)

$$\Phi(p) = \frac{G(p)}{1 + G(p)} = \frac{G_1(p)}{p^k + G_1(p)}$$

ou

$$\Phi(p) = \frac{1}{1 + \dfrac{p^k}{G_1(p)}}. \tag{5.78}$$

Pour des valeurs de p petites en module (c'est-à-dire dans le voisinage du point $p = 0$), la fonction $\Phi(p)$ peut être représentée par la série convergente suivante:

$$\Phi(p) = 1 - \frac{p^k}{G_1(p)} + \frac{p^{2k}}{G_1^2(p)} - \ldots \tag{5.79}$$

L'expression (79) implique

$$\Phi(0) = 1, \ \Phi'(0) = 0, \ \Phi''(0) = 0, \ \ldots, \ \Phi^{(k-1)}(0) = 0,$$
$$\Phi^{(k)}(0) = -\frac{k!}{G_1(0)} \neq 0. \tag{5.80}$$

D'après (80), les constantes d'erreur d'un système asservi, définies par les expressions (74), sont

$$c_0 = 0, \quad c_1 = 0, \ \ldots, \ c_{k-1} = 0, \ c_k \neq 0, \tag{5.81}$$

c'est-à-dire si la condition (76) est observée, le système asservi considéré est en effet un système astatique de degré k.

De cette façon, pour qu'un système asservi en boucle fermée soit un système astatique de degré k il faut que la fonction $G(p)$ possède en $p = 0$ un pôle d'ordre k.

CHAPITRE 2

SYSTÈMES COMMANDÉS NON LINÉAIRES

§ 6. Stabilité des systèmes commandés non linéaires. Critères fréquentiels. Application de la méthode directe de Liapounov

1. A propos d'une classe de systèmes commandés non linéaires. Nous allons examiner les oscillations libres d'un système non linéaire décrit par les équations différentielles suivantes :

$$\left.\begin{aligned} \frac{dx_j}{dt} &= \sum_{k=1}^{n} a_{jk}x_k + b_j y \quad (j=1, \ldots, n), \\ y &= \varphi(\sigma), \\ \sigma &= \sum_{k=1}^{n} c_k x_k, \end{aligned}\right\} \tag{6.1}$$

où la fonction $\varphi(\sigma)$ vérifie la condition

$$\varphi(0) = 0. \tag{6.2}$$

Le système d'équations (1) peut s'écrire :

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} a_{jk}x_k + b_j\varphi\left(\sum_{k=1}^{n} c_k x_k\right) \quad (j=1, \ldots, n). \tag{6.3}$$

Si l'on introduit les matrices

$$\left.\begin{aligned} a = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \ldots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \ldots \\ b_n \end{bmatrix}, \\ c = [c_1 \ldots c_n] \end{aligned}\right\} \tag{6.4}$$

et si l'on désigne par D l'opérateur de différentiation par rapport au temps

$$D = \frac{d}{dt},$$

on peut remplacer le système d'équations différentielles (3) par l'équation vectorielle

$$(ED - a)\,x - b\varphi(cx) = 0, \tag{6.5}$$

où par E on désigne la matrice unité.

En plus de la condition (2), on suppose que le graphique de la fonction $\varphi(\sigma)$ se trouve tout entier dans le domaine angulaire déli-

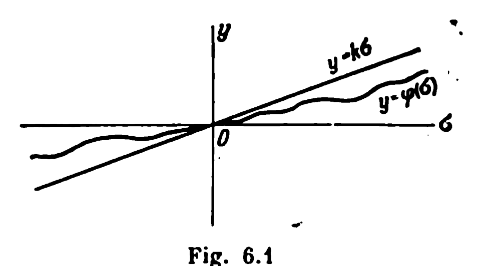


Fig. 6.1

mité par la figure 6.1, c'est-à-dire que

$$0 \leqslant \frac{\varphi(\sigma)}{\sigma} \leqslant k. \qquad (6.6)$$

Avec $k = \infty$, la condition (6) se ramène à l'inégalité

$$\sigma\varphi(\sigma) \geqslant 0.$$

Pour $b = 0$, l'équation vectorielle (5) se met sous la forme

$$(ED - a)\, x = 0. \qquad (6.7)$$

A l'équation différentielle (7) correspond l'équation caractéristique

$$\Delta(p) = 0, \qquad (6.8)$$

où

$$\Delta(p) = \det(Ep - a). \qquad (6.9)$$

Nous allons appeler *cas principal* celui où toutes les racines de l'équation caractéristique se situent dans le plan de la variable complexe p à gauche de l'axe imaginaire, c'est-à-dire

$$\mathrm{Re}\, p_j < 0 \quad (j = 1, \ldots, n).$$

Dans ce qui suit nous nous bornerons à l'étude du cas principal.

Le système d'équations (1) peut s'écrire sous la forme vectorielle

$$(ED - a)\, x = by, \quad y = \varphi(\sigma), \quad \sigma = cx. \qquad (6.10)$$

En éliminant x de l'équation (10) on obtient l'équation suivante :

$$\sigma = c\,(ED - a)^{-1}\, by. \qquad (6.11)$$

Introduisons maintenant les notations

$$W(D) = -c\,(ED - a)^{-1}\, b. \qquad (6.12)$$

L'équation (11) devient

$$\sigma = -W(D)\, y. \qquad (6.13)$$

6*

La matrice $W(D)$ peut se mettre sous la forme:

$$W(D) = -c\frac{\operatorname{adj}(ED-a)}{\det(ED-a)}b = \frac{M(D)}{\Delta(D)}. \tag{6.14}$$

L'expression (14) montre que la fonction $W(D)$ est une fraction rationnelle scalaire dont le degré du numérateur est inférieur à celui du dénominateur.

2. Interprétation de la fonction $W(D)$. A un système commandé dont les oscillations libres sont décrites par les équations (1), on peut

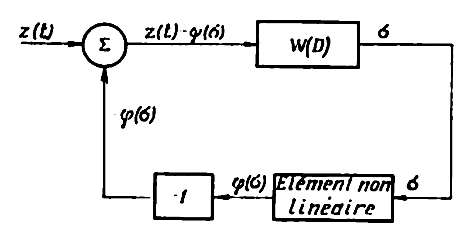


Fig. 6.2

faire correspondre le diagramme fonctionnel représenté sur la figure 6.2. Ce diagramme est celui d'un système en boucle fermée dont la chaîne de retour comporte un élément non linéaire. Par $z(t)$ on désigne le signal d'entrée. Le diagramme concerné est décrit par l'équation

$$\sigma = W(D)\,[z(t) - \varphi(\sigma)], \tag{6.15}$$

d'où l'on tire que

$$\sigma + W(D)\,\varphi(\sigma) = W(D)\,z(t). \tag{6.16}$$

Dans le cas particulier où

$$\varphi(\sigma) = h\sigma, \tag{6.17}$$

l'équation (16) devient:

$$[1 + hW(D)]\,\sigma = W(D)\,z(t), \tag{6.18}$$

ou

$$\sigma = \frac{W(D)}{1 + hW(D)}z(t). \tag{6.19}$$

Puisque d'après (14)

$$\frac{W(D)}{1 + hW(D)} = \frac{M(D)}{\Delta(D) + hM(D)}, \tag{6.20}$$

l'équation (19) peut s'écrire

$$[\Delta(D) + hM(D)]\,\sigma = M(D)\,z(t). \tag{6.21}$$

Les oscillations libres d'un système *en boucle fermée* pour $\varphi(\sigma) = h\sigma$ sont décrites par une équation homogène qui s'obtient à partir

de (21) avec $z(t) \equiv 0$:

$$[\Delta(D) + hM(D)]\,\sigma = 0. \qquad (6.22)$$

L'équation caractéristique qui correspond à l'équation différentielle (22) est

$$\Delta(p)\left[1 + h\frac{M(p)}{\Delta(p)}\right] = \Delta(p)\,[1 + hW(p)] = 0. \qquad (6.23)$$

Dans le cas principal examiné tous les zéros du polynôme $\Delta(p)$ se situent dans le demi-plan gauche de la variable complexe p. Par conséquent, pour que le système en boucle fermée soit asymptotiquement stable lorsque $\varphi(\sigma) = h\sigma$, c'est-à-dire que l'équation caractéristique (23) n'ait pas de racines dans le demi-plan droit p,

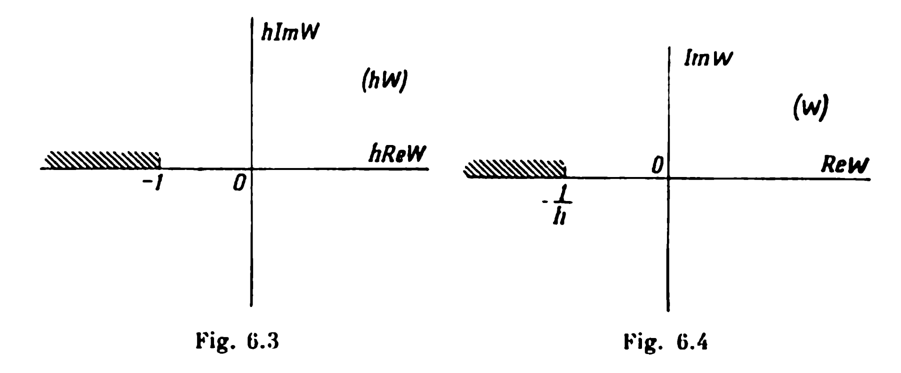


Fig. 6.3 Fig. 6.4

en vertu du critère de Nyquist, il faut et il suffit que lors de la rotation autour de l'hodographe du vecteur $hW(i\omega)$, le nombre de tours qu'on fait autour du point $(-1, 0)$ soit nul (fig. 6.3), ou que lors de la rotation de l'hodographe du vecteur $W(i\omega)$ le nombre de tours qu'on fait autour du point $\left(-\frac{1}{h}, 0\right)$ soit nul (fig. 6.4). Etant donné que les fonctions $\varphi(\sigma)$ vérifient la condition (6), les fonctions linéaires $h\sigma$ qui font partie de cette classe respectent la condition $0 < h \leqslant k$, ou

$$-\infty < -\frac{1}{h} \leqslant -\frac{1}{k}. \qquad (6.24)$$

Si l'hodographe du vecteur $W(i\omega)$ est de la forme représentée sur la figure 3.2, il existe une valeur $h = h^*$, vérifiant l'inégalité (24), telle que le point $-\frac{1}{h^*}$ se situe à l'intérieur de la branche suivant laquelle la rotation du rayon vecteur ayant pour origine le point $\left(-\frac{1}{h^*}, 0\right)$ a lieu dans le même sens que la rotation suivant la branche extérieure de l'hodographe. Par conséquent, pour $h = h^*$, le système

en boucle fermée est instable. Par suite pour qu'un système en boucle fermée soit asymptotiquement stable *quelle que soit la fonction* $\varphi(\sigma) = h\sigma$, où $h \in (0, k]$, il faut et il suffit que l'hodographe du vecteur $W(i\omega)$ ne coupe pas l'axe réel sur l'intervalle $\left(-\infty, -\frac{1}{k}\right]$ (fig. 6.5). Si $k = \infty$, on a $0 < h < \infty$, c'est-à-dire la zone interdite

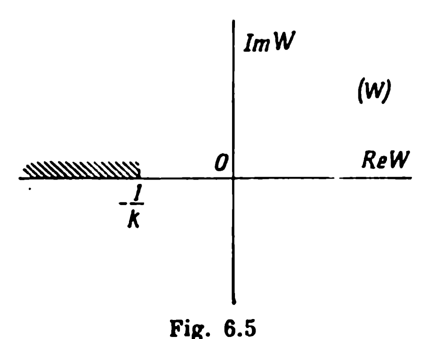


Fig. 6.5

est l'intervalle $(-\infty, 0)$. L'origine des coordonnées elle-même ne fait pas partie de la zone interdite, car nous examinons les fonctions $\varphi(\sigma) = h\sigma$ pour un h aussi grand que l'on veut mais fini.

3. Lieu de transfert modifié. Nous avons utilisé plus haut la relation (14) pour introduire la fonction

$$W(D) = \frac{M(D)}{\Delta(D)}.$$

Coupons la chaîne de retour du système représenté sur la figure 6.2. Le système en boucle ouverte de la figure 6.6 sera alors décrit par l'équation

$$\sigma = W(D)\, z(t), \tag{6.25}$$

où par $z(t)$ on désigne le signal appliqué à l'entrée du système.

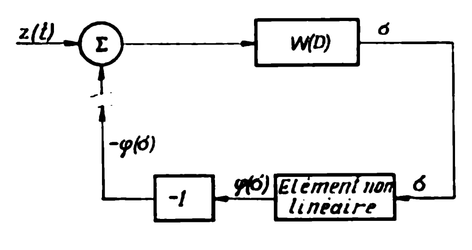


Fig. 6.6

Comme il résulte de l'équation (25), la fonction $W(D)$ est la fonction de transfert du système *en boucle ouverte* représentée sur la figure 6.6.

Conformément à l'expression (14), le lieu de transfert d'un système ouvert est

$$W(i\omega) = \frac{M(i\omega)}{\Delta(i\omega)}. \tag{6.26}$$

Introduisons maintenant la fonction $W^*(\omega)$ définie par les relations suivantes:

$$\left.\begin{aligned} \operatorname{Re} W^*(\omega) &= \operatorname{Re} W(i\omega), \\ \operatorname{Im} W^*(\omega) &= \omega \operatorname{Im} W(i\omega). \end{aligned}\right\} \tag{6.27}$$

Ainsi, la fonction $W^*(\omega)$ s'écrit

$$W^*(\omega) = \operatorname{Re} W(i\omega) + i\omega \operatorname{Im} W(i\omega). \tag{6.28}$$

Puisque $\operatorname{Re} W(i\omega)$ est une fonction paire de ω, et $\operatorname{Im} W(i\omega)$ est une fonction impaire de ω, en vertu de (27), la fonction $\operatorname{Im} W^*(\omega)$ est une fonction paire de ω et on a

$$W^*(-\omega) = W^*(\omega). \tag{6.29}$$

La fonction $W^*(\omega)$ s'appelle *lieu de transfert modifié*.

Puisqu'n vertu de (14) la fonction $W(D)$ est une fraction rationnelle dont le degré du numérateur est inférieur à celui du dénominateur, on aura pour $W(p)$ l'expression suivante

$$W(p) = \frac{M(p)}{\Delta(p)} = \frac{B_0 p^m + B_1 p^{m-1} + \ldots + B_{m-1} p + B_m}{p^n + A_1 p^{n-1} + \ldots + A_{n-1} p + A_n} \quad (m < n). \tag{6.30}$$

Dans le dénominateur de (30), le coefficient de p^n est égal à l'unité, car d'après (9)

$$\Delta(p) = \det(Ep - a).$$

En posant $p = i\omega$, on obtient d'après (30) que

$$W(i\omega) = \frac{M(i\omega)}{\Delta(i\omega)} = \frac{B_0 (i\omega)^m + B_1 (i\omega)^{m-1} + \ldots + B_{m-1} i\omega + B_m}{(i\omega)^n + A_1 (i\omega)^{n-1} + \ldots + A_{n-1} i\omega + A_n} \quad (m < n). \tag{6.31}$$

L'expression (31) entraîne

$$\lim_{\omega \to 0} W(i\omega) = \frac{B_m}{A_n}, \qquad \lim_{\omega \to \infty} W(i\omega) = 0. \tag{6.32}$$

En multipliant les premier et second membres de l'expression (31) par ω, on obtient

$$\omega W(i\omega) = -i \frac{i\omega M(i\omega)}{\Delta(i\omega)} =$$
$$= -i \frac{B_0 (i\omega)^{m+1} + B_1 (i\omega)^m + \ldots + B_{m-1} (i\omega)^2 + B_m i\omega}{(i\omega)^n + A_1 (i\omega)^{n-1} + \ldots + A_{n-1} i\omega + A_n} \quad (m < n). \tag{6.33}$$

En tenant compte que suivant (27)

$$\operatorname{Re} W^* (\omega) = \operatorname{Re} W (i\omega),$$
$$\operatorname{Im} W^* (\omega) = \omega \operatorname{Im} W (i\omega) = \operatorname{Im} [\omega W (i\omega)],$$

on trouve que

$$\lim_{\omega \to 0} \operatorname{Re} W^* (\omega) = \frac{B_m}{A_n}, \quad \lim_{\omega \to 0} \operatorname{Im} W^* (\omega) = 0,$$

et par conséquent,

$$\lim_{\omega \to 0} W^* (\omega) = \frac{B_m}{A_n}. \tag{6.34}$$

La valeur de $\lim\limits_{\omega \to \infty} W^* (\omega)$ diffère suivant que $n - m = 1$ ou $n - m > 1$.

Pour $n - m = 1$, c'est-à-dire pour $n = m + 1$, on a d'après (31) et (33)

$$\lim_{\omega \to \infty} \operatorname{Re} W^* (\omega) = 0, \quad \lim_{\omega \to \infty} \operatorname{Im} W^* (\omega) = -B_0.$$

Pour $n - m > 1$, c'est-à-dire pour $n = m + 2, m + 3, \ldots$, on a

$$\lim_{\omega \to \infty} \operatorname{Re} W^* (\omega) = 0, \quad \lim_{\omega \to \infty} \operatorname{Im} W^* (\omega) = 0.$$

Ainsi,

$$\lim_{\omega \to \infty} W^* (\omega) = \begin{cases} -iB_0 & \text{pour } n - m = 1, \\ 0 & \text{pour } n - m > 1. \end{cases} \tag{6.35}$$

4. Théorème de V.-M. Popov. Définition. *Le système* (1)

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} a_{jk} x_k + b_j y \qquad (j = 1, \ldots, n),$$
$$y = \varphi (\sigma),$$
$$\sigma = \sum_{k=1}^{n} c_k x_k,$$

dont la solution triviale $x_1 = x_2 = \ldots = x_n = 0$ *est asymptotiquement stable quelles que soient les formes de la fonction* $\varphi (\sigma)$ *respectant la condition* (6)

$$0 \leqslant \frac{\varphi (\sigma)}{\sigma} \leqslant k$$

est dit absolument stable dans l'angle $[0, k]$.

Comme les fonctions linéaires $\varphi (\sigma) = h\sigma$ $(0 < h \leqslant k)$ font partie des fonctions $\varphi (\sigma)$ admissibles d'après la condition (6), la condition imposée à l'hodographe du vecteur $W (i\omega)$ et qui consiste à ne pas couper l'axe réel sur l'intervalle $\left(-\infty, -\frac{1}{k}\right]$ (fig. 6.5) est une condition *nécessaire* de la stabilité absolue du système (1). La condition

suffisante de la stabilité absolue est donnée par le théorème de V. M. Popov.

Théorème de V.-M. Popov. *Pour rendre le système* (1) *absolument stable dans l'angle* $[0, k]$, *il suffit qu'il existe un nombre réel fini q tel que pour toute valeur de* $\omega \geqslant 0$ *la condition*

$$\operatorname{Re}(1+iq\omega)\,W(i\omega)+\frac{1}{k}>0 \qquad (6.36)$$

soit vérifiée.

5. Formulation géométrique du théorème de Popov. Notons que la condition *suffisante* (36) de la stabilité absolue d'un système non linéaire formulée dans le théorème de Popov se distingue nettement des contraintes imposées aux systèmes linéaires par le critère de Nyquist. Ce dernier n'impose une limitation de la valeur de $\operatorname{Re} W(i\omega)$ qu'aux points où $\operatorname{Im} W(i\omega) = 0$. En d'autres points la valeur de $\operatorname{Re} W(i\omega)$ peut être quelconque du fait que la limitation de la valeur de $\operatorname{Re} W(i\omega)$ aux points où $\operatorname{Im} W(i\omega) = 0$ rend impossible l'encerclement du point $\left(-\frac{1}{k},\ i0\right)$ par l'hodographe du vecteur $W(i\omega)$, ce qui assure précisément la stabilité asymptotique du système linéaire en boucle fermée. Quant à la condition suffisante (36) de la stabilité absolue d'un système non linéaire, formulée dans le théorème de Popov, elle impose la limitation des valeurs de $\operatorname{Re}(1+iq\omega)\ W(i\omega)$ pour tout $\omega \geqslant 0$, et non seulement aux points où $\operatorname{Im} W(i\omega) = 0$.

Nous avons introduit précédemment le lieu de transfert modifié $W^*(\omega)$. Désignons par X et Y respectivement les parties réelle et imaginaire de $W^*(\omega)$. D'après (27), on a

$$\left.\begin{aligned} X &= \operatorname{Re} W^*(\omega) = \operatorname{Re} W(i\omega), \\ Y &= \operatorname{Im} W^*(\omega) = \omega \operatorname{Im} W(i\omega). \end{aligned}\right\} \qquad (6.37)$$

En vertu de (37)

$$\operatorname{Re}(1+iq\omega)\ W(i\omega) = \\ = \operatorname{Re} W(i\omega) - q\omega \operatorname{Im} W(i\omega) = X - qY. \qquad (6.38)$$

Ainsi, la condition (36) devient

$$X-qY+\frac{1}{k}>0 \quad (\text{pour tout } \omega\geqslant 0). \qquad (6.39)$$

Dans le plan XY (c'est-à-dire dans le plan de la variable complexe W^*) l'équation

$$X-qY+\frac{1}{k}=0,$$

ou l'équation équivalente

$$Y=\frac{1}{q}\left(X+\frac{1}{k}\right), \qquad (6.40)$$

détermine une droite (fig. 6.7) qui passe par le point $(-1/k, 0)$. La pente de cette droite est égale à $1/q$. La droite (40) peut être nommée droite de Popov.

On voit sans peine que la condition (39) est vérifiée en tout point du plan de la variable complexe W^*, qui repose à droite de la droite de Popov. Autrement dit, la condition (39) traduit le fait que l'hodographe du vecteur $W^*(\omega)$ doit se situer à droite de la droite de Popov.

Formulation géométrique du théorème de Popov. *Pour que le système* (1) *soit absolument stable dans l'angle* $[0, k]$, *il suffit qu'on puisse choisir dans le plan de la variable*

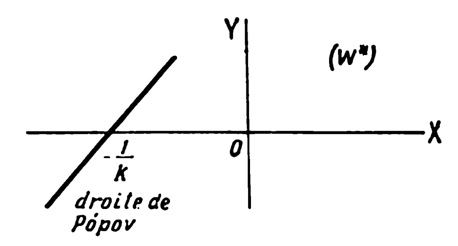


Fig. 6.7

complexe W^* *une droite qui passe par le point* $(-1/k, i0)$ *tel que l'hodographe du vecteur* $W^*(\omega)$ *se situe tout entier à droite de cette ligne.*

6. Sur la possibilité de démontrer le théorème en se bornant au cas $q > 0$. Adressons-nous au système d'équations (1). Transformons ce système en posant

$$y = k\sigma - \tilde{y}. \qquad (6.41)$$

Comme d'après (1) $y = \varphi(\sigma)$, la relation (41) peut s'écrire:

$$\varphi(\sigma) = k\sigma - \tilde{\varphi}(\sigma). \qquad (6.42)$$

Les fonctions $\varphi(\sigma)$ satisfont à la condition (6), c'est-à-dire les courbes des fonctions $\varphi(\sigma)$ se situent dans l'angle $[0, k]$ (fig. 6.1). Des relations (42) et (6) il ressort que les fonctions $\tilde{\varphi}(\sigma)$ également respectent la condition

$$0 \leqslant \frac{\tilde{\varphi}(\sigma)}{\sigma} \leqslant k,$$

c'est-à-dire que les courbes des fonctions $\tilde{\varphi}(\sigma)$ ne sortent pas non plus de l'angle $[0, k]$. Ainsi, si le système initial (1) est absolument stable dans l'angle $[0, k]$, le système transformé à l'aide de la relation (41) est également absolument stable dans cet angle.

D'après (13)

$$-\sigma = W(D)\, y.$$

En remplaçant y par sa valeur tirée de (41), on obtient

$$-\sigma = W(D)(k\sigma - \widetilde{y}). \tag{6.43}$$

Il en résulte que

$$-\sigma = \widetilde{W}(D)\widetilde{y}, \tag{6.44}$$

où

$$\widetilde{W}(D) = -\frac{W(D)}{1+kW(D)}. \tag{6.45}$$

La relation (45) entraîne

$$\widetilde{W}(i\omega) = -\frac{W(i\omega)}{1+kW(i\omega)}. \tag{6.46}$$

On vérifie aisément l'identité suivante:

$$\operatorname{Re}(1-iq\omega)\widetilde{W}(i\omega)+\frac{1}{k} = \frac{\operatorname{Re}(1+iq\omega)W(i\omega)+\frac{1}{k}}{|1+kW(i\omega)|^2}. \tag{6.47}$$

En effet,

$$\operatorname{Re}(1-iq\omega)\widetilde{W}+\frac{1}{k} = \operatorname{Re}\left[(-1+iq\omega)\frac{W}{1+kW}+\frac{1}{k}\right] =$$

$$= \operatorname{Re}\frac{-W+iq\omega W+\frac{1}{k}+W}{1+kW} = \frac{1}{|1+kW|^2}\operatorname{Re}\left[\left(iq\omega W+\frac{1}{k}\right)(1+k\overline{W})\right] =$$

$$= \frac{1}{|1+kW|^2}\operatorname{Re}\left[iq\omega W+\frac{1}{k}+\overline{W}+iq\omega k|W|^2\right] = \frac{\operatorname{Re}\left[(1+iq\omega)W+\frac{1}{k}\right]}{|1+kW|^2},$$

ce qu'il fallait démontrer.

Etant donné que tous les pôles de la fonction $W(p) = \frac{M(p)}{\Delta(p)}$ reposent à gauche de l'axe imaginaire, la valeur de $W(i\omega)$ est finie quelle que soit la valeur de ω. Par suite si le système considéré satisfait à la relation

$$\operatorname{Re}(1+iq\omega)W(i\omega)+\frac{1}{k} > 0, \tag{6.48}$$

en vertu de (47) on a également la relation suivante

$$\operatorname{Re}(1+i\widetilde{q}\omega)\widetilde{W}(i\omega)+\frac{1}{k} > 0, \tag{6.49}$$

où

$$\widetilde{q} = -q < 0. \tag{6.50}$$

Nous avons dit plus haut que si le système initial est absolument stable dans l'angle $[0, k]$, le système transformé le sera également. Mais pour le système transformé, la condition du théorème de Popov est exprimée par l'inégalité (49) où $\widetilde{q} = -q < 0$.

D'après ce qui vient d'être dit pour démontrer le théorème de Popov, on peut se borner à l'examen du cas $q > 0$. L'étude du cas $\tilde{q} = -q < 0$ ne peut pas étendre la classe des systèmes soumis à la condition suffisante de Popov, du fait qu'avec $\tilde{q} = -q$ on obtient le même système (à stabilité déjà établie par la relation (36), où $q > 0$), mais ramené seulement à une nouvelle variable $\tilde{y}$.

7. Lemme 1. Soient

1) $f(t)$ une fonction continue bornée pour $t \geqslant 0$ avec sa dérivée $\frac{df}{dt}$;

2) $G(x) > 0$ une fonction continue pour tout $x \neq 0$ et telle que $G(0) = 0$;

3) $\int\limits_0^\infty G[f(t)]\,dt < \infty$.

Montrer que dans ces conditions

$$\lim_{t\to\infty} f(t) = 0.$$

D é m o n s t r a t i o n. Soit pour $t > 0$

$$|f(t)| < a < \infty, \quad \left|\frac{df}{dt}\right| < b < \infty, \tag{6.51}$$

$$\int\limits_0^\infty G[f(t)]\,dt = c < \infty. \tag{6.52}$$

Supposons que l'affirmation du lemme soit fausse. La définition de la limite entraîne alors qu'il existe une suite croissante illimitée de nombres $t_k > 0$ telle que

$$|f(t_k)| \geqslant \varepsilon > 0 \quad (k = 1, 2, \ldots). \tag{6.53}$$

De plus, on peut toujours considérer que pour tout k

$$t_{k+1} - t_k \geqslant m > 0, \tag{6.54}$$

car si cette condition n'est pas observée pour la suite initiale, en omettant une partie de termes de la suite on peut parvenir à faire respecter cette inégalité.

Puisque $\left|\frac{df}{dt}\right| < b$, on a pour tout $t \geqslant 0$

$$|f(t) - f(t_k)| \leqslant b\,|t - t_k|. \tag{6.55}$$

Compte tenu du fait que $G(x) > 0$ pour tout $x \neq 0$, on a

$$\int\limits_0^\infty G[f(t)]\,dt \geqslant \sum_{k=1}^\infty \int\limits_{t_k-\frac{m}{2}}^{t_k+\frac{m}{2}} G[f(t)]\,dt. \tag{6.56}$$

Mais toutes les valeurs de t telles que

$$t_k - \frac{m}{2} \leqslant t \leqslant t_k + \frac{m}{2}$$

donnent lieu à l'inégalité

$$|f(t)| = |f(t_k) + [f(t) - f(t_k)]| \geqslant$$
$$\geqslant |f(t_k)| - |f(t) - f(t_k)| \geqslant \varepsilon - b\frac{m}{2} = \varepsilon_0. \qquad (6.57)$$

Si l'on choisit m suffisamment petit, ε_0 sera positif

$$\varepsilon_0 > 0. \qquad (6.58)$$

Si l'on désigne

$$r = \min_{\varepsilon_0 \leqslant |x| \leqslant a} G(x), \qquad (6.59)$$

où, en vertu du lemme, $r > 0$, on a l'estimation

$$\int_{t_k - \frac{m}{2}}^{t_k + \frac{m}{2}} G[f(t)]\,dt \geqslant rm. \qquad (6.60)$$

On en déduit, en vertu de l'inégalité (56),

$$\int_0^\infty G[f(t)]\,dt = \infty, \qquad (6.61)$$

ce qui ramène à l'absurde la condition du lemme. Cette contradiction prouve le lemme.

8. Lemme 2. Si avec $t \to \infty$ les trois fonctions réelles $f_1(t)$, $f_2(t)$, $f_3(t)$ tendent vers zéro aussi ou plus vite que les exponentielles (ce que nous noterons $f_h(t) \rightrightarrows 0$) et si leurs transformées de Fourier

$$F_k(i\omega) = \int_0^\infty f_k(t)\,e^{-i\omega t}\,dt \qquad (k = 1, 2, 3) \qquad (6.62)$$

sont liées par la relation

$$F_1(i\omega) = H(i\omega)\,F_3(i\omega) + F_2(i\omega), \qquad (6.63)$$

où

$$\operatorname{Re} H(i\omega) \geqslant \delta > 0 \qquad (6.64)$$

pour tout ω, alors

$$-\int_0^\infty f_1(t)\,f_3(t)\,dt \leqslant C \qquad (C > 0), \qquad (6.65)$$

où la valeur de la constante positive C est finie et où cette constante se détermine par l'égalité

$$C=\frac{1}{8\pi\delta}\int_{-\infty}^{+\infty}|F_2(i\omega)|^2\,d\omega. \qquad (6.66)$$

Démonstration. Puisque $f_1(t) \rightrightarrows 0$, $f_3(t) \rightrightarrows 0$, on peut appliquer la formule de Parseval [17, 90]

$$\int_0^\infty f_1(t)\,f_3(t)\,dt=\frac{1}{2\pi}\int_{-\infty}^\infty \overline{F_1(i\omega)}\,F_3(i\omega)\,d\omega. \qquad (6.67)$$

En remplaçant $F_1(i\omega)$ par sa valeur tirée de (63), on obtient

$$\int_0^\infty f_1(t)\,f_3(t)\,dt=\frac{1}{2\pi}\int_{-\infty}^\infty \overline{[H(i\omega)\,F_3(i\omega)+F_2(i\omega)]}\,F_3(i\omega)\,d\omega. \qquad (6.68)$$

Le premier membre de la relation (68) étant réel, la partie imaginaire de l'intégrale du second membre de (68) est nulle. Il vient donc

$$\int_0^\infty f_1(t)\,f_3(t)\,dt=\frac{1}{2\pi}\int_{-\infty}^\infty \operatorname{Re}[\overline{H(i\omega)}\,|F_3(i\omega)|^2+\overline{F_2(i\omega)}\,F_3(i\omega)]\,d\omega=$$

$$=\frac{1}{2\pi}\int_{-\infty}^\infty \Big\{\operatorname{Re}H(i\omega)\,|F_3(i\omega)|^2+\frac{1}{2}[\overline{F_2(i\omega)}\,F_3(i\omega)+$$

$$+F_2(i\omega)\,\overline{F_3(i\omega)}]\Big\}\,d\omega=$$

$$=\frac{1}{2\pi}\int_{-\infty}^\infty \left[\sqrt{\operatorname{Re}H(i\omega)}\,F_3(i\omega)+\frac{F_2(i\omega)}{2\sqrt{\operatorname{Re}H(i\omega)}}\right]\times$$

$$\times\left[\sqrt{\operatorname{Re}H(i\omega)}\,\overline{F_3(i\omega)}+\frac{\overline{F_2(i\omega)}}{2\sqrt{\operatorname{Re}H(i\omega)}}\right]d\omega-$$

$$-\frac{1}{8\pi}\int_{-\infty}^\infty \frac{|F_2(i\omega)|^2}{\operatorname{Re}H(i\omega)}\,d\omega=\frac{1}{2\pi}\int_{-\infty}^\infty \Big|\sqrt{\operatorname{Re}H(i\omega)}\,F_3(i\omega)+$$

$$+\frac{F_2(i\omega)}{2\sqrt{\operatorname{Re}H(i\omega)}}\Big|^2\,d\omega-\frac{1}{8\pi}\int_{-\infty}^\infty \frac{|F_2(i\omega)|^2}{\operatorname{Re}H(i\omega)}\,d\omega. \qquad (6.69)$$

En rejetant le premier terme du second membre de la relation (69) qui est toujours positif et en tenant compte du fait que d'après (64)

$$\operatorname{Re}H(i\omega)\geqslant\delta>0,$$

on obtient que

$$\int_0^\infty f_1(t) f_3(t)\,dt \geqslant -\frac{1}{8\pi\delta}\int_{-\infty}^{\infty} |F_2(i\omega)|^2\,d\omega, \tag{6.70}$$

c'est-à-dire

$$-\int_0^\infty f_1(t) f_3(t)\,dt \leqslant \frac{1}{8\pi\delta}\int_{-\infty}^{\infty} |F_2(i\omega)|^2\,d\omega. \tag{6.71}$$

Comme d'après la condition du lemme, pour $t \to \infty$ la fonction $f_2(t)$ tend vers zéro aussi ou plus vite que l'exponentielle, c'est-à-dire comme $\int_0^\infty f_2^2(t)\,dt$ converge, et suivant la formule de Parseval,

$$\int_0^\infty f_2^2(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |F_2(i\omega)|^2\,d\omega,$$

l'expression (71) se met sous la forme

$$-\int_0^\infty f_1(t) f_3(t)\,dt \leqslant C < \infty, \tag{6.72}$$

ce qu'il fallait démontrer.

9. **Démonstration du théorème de Popov.** En plus des articles de Popov [73], la méthode qu'il a développée et sa mise au point ultérieure ont donné lieu à une littérature abondante [3, 22, 54]. Parmi les nombreuses démonstrations de son théorème nous avons retenu celle donnée par la monographie [3].

Comme nous l'avons montré au point 6, il suffit de prouver le théorème sous l'hypothèse que dans la formule (36) $q > 0$. Soit $x_j(t)$ ($j = 1, \ldots, n$) une solution du système d'équations différentielles (1)

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} a_{jk}x_k + b_j y \qquad (j = 1, \ldots, n),$$
$$y = \varphi(\sigma),$$
$$\sigma = \sum_{k=1}^{n} c_k x_k$$

avec la fonction $\varphi(\sigma)$ choisie arbitrairement mais *fixée* dans l'angle $[0, k]$ et certaines conditions initiales également arbitraires mais fixées.

Si l'on porte les fonctions $x_j(t)$, qui représentent la solution indiquée plus haut, dans les formules

$$\sigma = \sum_{k=1}^{n} c_k x_k \quad \text{et} \quad y = \varphi(\sigma),$$

on obtient que pour le système (1) avec la fonction $\varphi(\sigma)$ fixée et les conditions initiales choisies,

$$\sigma = \sigma(t) = \sum_{k=1}^{n} c_k x_k(t), \tag{6.73}$$

$$y = y(t) = \varphi[\sigma(t)], \tag{6.74}$$

c'est-à-dire il se trouve que la fonction $y(t)$ est une fonction du temps *connue*.

Les fonctions $x_j(t)$ ($j = 1, \ldots, n$) indiquées plus haut peuvent donc être considérées comme solution du système d'équations *linéaires inhomogènes*

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} a_{jk} x_k + b_j y(t) \quad (j = 1, \ldots, n), \tag{6.75}$$

où la fonction $y(t)$ est déterminée par l'expression (74), alors que les conditions initiales sont les mêmes que celles données pour la résolution du système d'équations (1).

Adoptons maintenant la notation $y_T(t)$ pour la fonction suivante

$$y_T(t) = \begin{cases} y(t) & \text{pour} \quad t \leqslant T, \\ 0 & \text{pour} \quad t > T, \end{cases} \tag{6.76}$$

où T est un nombre positif fixe arbitraire, et soit la fonction

$$\sigma_T(t) = \sum_{k=1}^{n} c_k x_{l,T}(t) \tag{6.77}$$

déterminée par un système d'équations linéaires inhomogènes

$$\frac{dx_{jT}}{dt} = \sum_{k=1}^{n} a_{jk} x_{kT} + b_j y_T(t) \quad (j = 1, \ldots, n) \tag{6.78}$$

sous les mêmes conditions initiales suivant lesquelles a été déterminée d'après (73) la fonction $\sigma(t)$, c'est-à-dire

$$x_{jT}(0) = x_j(0) \quad (j = 1, \ldots, n). \tag{6.79}$$

Il en résulte que

$$\sigma_T(t) = \sigma(t) \quad \text{pour} \quad 0 \leqslant t \leqslant T. \tag{6.80}$$

La solution de l'équation différentielle vectorielle inhomogène

$$\frac{dx_T}{dt} = ax_T + by_T(t), \tag{6.81}$$

équivalente au système d'équations scalaires (78), peut s'écrire

$$x_T(t) = \theta(t)\,\theta^{-1}(0)\,x(0) + \theta(t)\int_0^t \theta^{-1}(\tau)\,by_T(\tau)\,d\tau, \qquad (6.82)$$

où $\theta(t)$ est la matrice fondamentale des solutions de l'équation vectorielle homogène

$$\frac{dx_T}{dt} = ax_T. \qquad (6.83)$$

Les éléments $\theta_{jk}(t)$ de la matrice $\theta(t)$ sont de la forme:

$$\theta_{jk}(t) = P_{jk}(t)\,e^{p_k t}, \qquad (6.84)$$

où $P_{jk}(t)$ sont des polynômes en t, et p_k, les zéros du déterminant caractéristique (9)

$$\Delta(p) = \det(Ep - a).$$

Dans le cas principal qui nous intéresse, tous les zéros du déterminant caractéristique $\Delta(p)$ possèdent une partie réelle négative $\operatorname{Re} p_k < 0$ et, par suite,

$$\lim_{t\to\infty} \theta_{jk}(t) = 0 \quad (j,\ k = 1, \ldots, n). \qquad (6.85)$$

Il résulte de (82), (77) et (4) que

$$\sigma_T(t) = \tilde{\sigma}_T(t) + \rho(t), \qquad (6.86)$$

où

$$\tilde{\sigma}_T(t) = c\theta(t)\int_0^t \theta^{-1}(\tau)\,by_T(\tau)\,d\tau, \qquad (6.87)$$

$$\rho(t) = c\theta(t)\,\theta^{-1}(0)\,x(0). \qquad (6.88)$$

De la sorte, $\tilde{\sigma}_T(t)$ est déterminée à partir de la solution (82) du système d'équations (78) sous les conditions initiales nulles, alors que $\rho(t)$ est la fonction du temps linéairement dépendante des valeurs initiales $x_j(0)$ $(j = 1, \ldots, n)$. Il s'ensuit de (85) que dans le cas principal étudié ici

$$\rho(t) \rightrightarrows 0, \quad \frac{d\rho(t)}{dt} \rightrightarrows 0. \qquad (6.89)$$

En dérivant (86), on a

$$\dot{\sigma}_T(t) = \dot{\tilde{\sigma}}_T(t) + \dot{\rho}(t). \qquad (6.90)$$

(86) et (90) permettent d'obtenir l'expression suivante:

$$\sigma_T(t) + q\dot{\sigma}_T(t) - \frac{y_T(t)}{k} = \tilde{\sigma}_T(t) + q\dot{\tilde{\sigma}}_T(t) - \frac{y_T(t)}{k} + \rho(t) + q\dot{\rho}(t). \quad (6.91)$$

Introduisons maintenant les notations

$$f_1(t) = -\sigma_T(t) - q\dot{\sigma}_T(t) + \frac{y_T(t)}{k}, \tag{6.92}$$

$$f_2(t) = -\rho(t) - q\dot{\rho}(t).$$

Désignons respectivement les transformées de Fourier des fonctions $f_1(t), f_2(t), y_T(t), \tilde{\sigma}_T(t)$ par $F_1(i\omega)$, $F_2(i\omega)$, $Y_T(i\omega)$, $\tilde{\Sigma}_T(i\omega)$. Ces transformations de Fourier existent du fait que $y_T(t) \equiv 0$ pour $t > T$, c'est-à-dire $y_T(t) \rightrightarrows 0$. La fonction $\tilde{\sigma}_T(t)$ est déterminée par l'expression (87), dans laquelle la fonction sous le signe somme s'annule avec $t > T$; conformément à (85), on en tire que $\tilde{\sigma}_T(t) \rightrightarrows$

$\rightrightarrows 0$, $\dot{\tilde{\sigma}}_T(t) \rightrightarrows 0$. En outre d'après (88) et (85), $\rho(t) \rightrightarrows 0$,

$\dot{\rho}(t) \rightrightarrows 0$. Ainsi, $f_1(t) \rightrightarrows 0$, $f_2(t) \rightrightarrows 0$.

D'après (13)

$$-\sigma(t) = W(D)\, y(t).$$

(14) et (78) impliquent

$$-\sigma_T(t) = W(D)\, y_T(t). \tag{6.93}$$

Dans les conditions initiales nulles, à l'équation (93) correspond la transformée de Fourier suivante:

$$-\tilde{\Sigma}_T(i\omega) = W(i\omega)\, Y_T(i\omega). \tag{6.94}$$

Ici à gauche figure $-\tilde{\Sigma}_T(i\omega)$, car il s'agit de la résolution sous les conditions initiales nulles, c'est-à-dire $\tilde{\sigma}_T(t)$. Dans les nouvelles notations (92), la relation (91) devient

$$f_1(t) = -\tilde{\sigma}_T(t) - q\dot{\tilde{\sigma}}_T(t) + \frac{y_T(t)}{k} + f_2(t). \tag{6.95}$$

En passant aux transformées de Fourier et en tenant compte de (94), on obtient

$$F_1(i\omega) = \left[(1 + iq\omega)\, W(i\omega) + \frac{1}{k}\right] Y_T(i\omega) + F_2(i\omega). \tag{6.96}$$

Ici on a tenu compte du fait que la transformée de Fourier de $\dot{\tilde{\sigma}}_T(t)$ est $i\omega\tilde{\Sigma}_T(i\omega)$. Désignons maintenant

$$H(i\omega) = (1 + iq\omega)\, W(i\omega) + \frac{1}{k}. \tag{6.97}$$

La relation (96) devient alors

$$F_1(i\omega) = H(i\omega)\,Y_T(i\omega) + F_2(i\omega). \qquad (6.98)$$

Dans notre cas *principal* la condition (36) du théorème de Popov peut être remplacée par l'inégalité

$$\operatorname{Re} H(i\omega) \geqslant \delta, \qquad (6.99)$$

où δ est un nombre positif. Comme il a été montré au point 3, dans le cas principal l'hodographe du vecteur $W^*(\omega)$ repose toujours dans une partie bornée du plan de la variable complexe W^* pour tout ω. La distance minimale entre les points de la droite de Popov (40) et l'hodographe du vecteur $W^*(\omega)$ est un nombre positif, c'est-à-dire les coordonnées X et Y d'un point quelconque de l'hodographe du vecteur $W^*(\omega)$ vérifient la condition

$$X - qY + \frac{1}{k} \geqslant \delta > 0, \qquad (6.100)$$

d'où l'on déduit précisément la condition (99).

De plus, comme nous l'avons établi précédemment,

$$f_1(t) \rightrightarrows 0, \qquad f_2(t) \rightrightarrows 0, \qquad y_T(t) \rightrightarrows 0.$$

Donc, d'après le lemme 2, on a

$$-\int_0^\infty f_1(t)\,y_T(t)\,dt \leqslant C, \qquad (6.101)$$

avec

$$C = \frac{1}{8\pi\delta} \int_{-\infty}^{\infty} |F_2(i\omega)|^2\,d\omega. \qquad (6.102)$$

Comme le montrent (88) et (92), la fonction $F_2(i\omega)$, de même que son original $f_2(t)$, dépend de façon linéaire et homogène des données initiales et ne dépend pas de la valeur de T. Par suite, la *constante positive* C dépend de façon quadratique et homogène des données initiales et tend vers zéro lorsque ces données initiales tendent vers zéro. *La constante C ne dépend pas de la grandeur T.*

En portant l'expression (92) de $f_1(t)$ dans (101), on obtient

$$\int_0^\infty y_T(t)\left[\sigma_T(t) + q\dot{\sigma}_T(t) - \frac{y_T(t)}{k}\right] dt =$$

$$= \int_0^T \varphi[\sigma(t)]\left\{\sigma(t) + q\dot{\sigma}(t) - \frac{\varphi[\sigma(t)]}{k}\right\} dt \leqslant C. \qquad (6.103)$$

En divisant le premier membre de (103) en deux intégrales, on a

$$\int_0^T \varphi[\sigma(t)]\left\{\sigma(t)-\frac{\varphi[\sigma(t)]}{k}\right\}dt+q\int_{\sigma(0)}^{\sigma(T)}\varphi(\sigma)\,d\sigma\leqslant C. \quad (6.104)$$

Ajoutons aux deux membres de l'inégalité (104) la grandeur positive $q\int_0^{\sigma(0)}\varphi(\sigma)\,d\sigma$ et introduisons une nouvelle constante

$$C^*=C+q\int_0^{\sigma(0)}\varphi(\sigma)\,d\sigma \quad (6.105)$$

pour aboutir à l'inégalité

$$\int_0^T \varphi(\sigma)\left[\sigma-\frac{\varphi(\sigma)}{k}\right]dt+q\int_0^{\sigma(T)}\varphi(\sigma)\,d\sigma\leqslant C^*. \quad (6.106)$$

Ici la constante C^*, de même que C, ne dépend que des données initiales et tend vers zéro avec ces dernières.

Chaque intégrale du premier membre de (106) est non négative. C'est pourquoi on a les inégalités

$$\int_0^T \varphi(\sigma)\left[\sigma-\frac{\varphi(\sigma)}{k}\right]dt\leqslant C^*, \quad (6.107)$$

$$\int_0^{\sigma(T)}\varphi(\sigma)\,d\sigma\leqslant\frac{1}{q}C^*. \quad (6.108)$$

Si $\varphi(\sigma)$ est comprise dans l'angle $[0, k]$, l'inégalité (108) n'entraîne pas encore que la fonction $\sigma(t)$ est bornée, puisqu'on peut choisir dans cet angle des caractéristiques de $\varphi(\sigma)$ telles que $\int_0^\infty \varphi(\sigma)\,d\sigma$ sera finie (par exemple, les fonctions $\varphi(\sigma)$ respectant la condition $\varphi(\sigma)\rightrightarrows 0$).

C'est pourquoi admettons d'abord que $\varphi(\sigma)$ satisfait à une condition plus serrée

$$0<\varepsilon\leqslant\frac{\varphi(\sigma)}{\sigma}<k, \quad (6.109)$$

que la condition principale (6)

$$0\leqslant\frac{\varphi(\sigma)}{\sigma}\leqslant k,$$

c'est-à-dire que $\varphi(\sigma)$ se situe dans l'angle $[\varepsilon, k)$ et non pas dans l'angle $[0, k]$. Alors l'inégalité (108) entraîne immédiatement que

la fonction $\sigma(t)$ est bornée pour $t \geqslant 0$, ou

$$|\sigma(t)| \leqslant L \quad (t \geqslant 0), \tag{6.110}$$

avec la grandeur L qui tend vers zéro avec les écarts initiaux, puisque C^* jouit de cette propriété.

Ainsi, par exemple, en remplaçant $\varphi(\sigma)$ par $\varepsilon\sigma$ dans le premier membre de (108), on obtient

$$\frac{1}{2}\varepsilon\sigma^2(T) \leqslant \frac{C^*}{q}. \tag{6.111}$$

D'où en remplaçant la grandeur arbitraire T par t, on amène l'inégalité (110) telle que

$$L = \sqrt{\frac{2C^*}{\varepsilon q}}. \tag{6.112}$$

Mais alors, en raison des propriétés connues des équations différentielles linéaires, dans le cas principal considéré les solutions $x_i(t)$ du système (75) sont également bornées

$$|x_i(t)| \leqslant N_i, \quad t \geqslant 0 \ (i = 1, \ldots, n), \tag{6.113}$$

où les constantes N_i, de même que la constante L, tendent vers zéro avec les écarts initiaux, du fait que les seconds membres de l'équation (75), $b_j y(t) = b_j \varphi[\sigma(t)]$ sont bornés si σ est bornée.

Ceci résulte de la propriété connue des solutions des systèmes d'équations linéaires inhomogènes relative au cas *stable* : si les seconds membres sont bornés dans l'intervalle $0 \leqslant t \leqslant \infty$ ou s'ils tendent vers zéro quand $t \to \infty$, toute solution jouit également de cette propriété, c'est-à-dire est respectivement bornée ou tend vers zéro quand $t \to \infty$. Si, de plus, les constantes qui délimitent les seconds membres, dépendent des écarts initiaux (comme c'est le fait dans notre cas) et tendent vers zéro avec ces écarts, toute solution jouit également de cette propriété.

De (113) il suit que la solution triviale du système (1) $x_i = 0$ est stable au sens de Liapounov. Pourtant, (113) n'entraîne pas encore la stabilité asymptotique de la solution triviale.

Adressons-nous maintenant à l'inégalité (107). Nous avons ici sous le signe somme la fonction

$$G(\sigma) = \varphi(\sigma)\left[\sigma - \frac{\varphi(\sigma)}{k}\right], \tag{6.114}$$

positive en vertu de l'inégalité (109) quelle que soit $\sigma \neq 0$ et $G(0) = 0$. D'autre part, le fait que les fonctions $x_i(t)$ sont bornées entraîne que d'après les équations (1) les dérivées $\dot{x}_i$ le sont également, de même que $\sigma(t)$ et $\dot{\sigma}(t)$. Remarquons que puisque d'après (102) et (105) la constante C^* ne dépend pas de T. l'inégalité (107) a lieu

pour $T > 0$ arbitraire. Il en résulte que

$$\int_0^\infty \varphi(\sigma)\left[\sigma - \frac{\varphi(\sigma)}{k}\right] dt < \infty.$$

Ainsi, toutes les conditions du lemme 1 sont remplies. Conformément à ce lemme, on conclut que sous des *conditions initiales arbitraires*

$$\lim_{t\to\infty} \sigma(t) = 0. \tag{6.115}$$

Il résulte de (115) que

$$\lim_{t\to\infty} \varphi[\sigma(t)] = 0. \tag{6.116}$$

La relation (116) peut s'écrire: $\lim\limits_{t\to\infty} y(t) = 0$. En outre, d'après la remarque sur les propriétés des solutions des équations différentielles (p. 101), il suit des équations (75) que

$$\lim_{t\to\infty} x_i(t) = 0 \qquad (i = 1, \ldots, n). \tag{6.117}$$

Ce qui vient d'être dit aurait mis fin à la démonstration du théorème de Popov si nous n'avions pas remplacé l'angle $[0, k]$ par l'angle $[\varepsilon, k)$. Cette restriction peut être levée. Toutefois, compte tenu du fait que la condition (109) permet de choisir ε aussi petit que l'on veut, nous n'allons pas nous y attarder.

10. Application de la méthode directe de Liapounov. Méthode de A. Lourié dans la théorie de la stabilité absolue des systèmes non linéaires. Le problème des conditions suffisantes de la stabilité absolue des systèmes décrits par les équations de la forme (1) a été pour la première fois posé et étudié par A. Lourié [59] à l'aide de la méthode de Liapounov. A cet effet, Lourié a proposé d'utiliser la fonction de Liapounov du type

$$V = L(x) + \vartheta \int_0^\sigma \varphi(\xi)\, d\xi. \tag{6.118}$$

Ici

$$L(x) = x^* H x, \tag{6.119}$$

où x^* est la matrice ligne obtenue par transposition du vecteur x, et H, la matrice carrée $n \times n$ aux éléments constants. La fonction $L(x)$ est une forme quadratique des variables x_j $(j = 1, \ldots, n)$. S'il est possible de choisir la matrice H de façon que $L(x)$ soit une forme quadratique positive, alors que la dérivée dV/dt de la fonction de Liapounov (118), existant grâce aux équations (1), soit une fonction négative de signe défini, le système considéré est asymptotiquement stable.

La relation entre les méthodes de Lourié et de Popov a fait l'objet d'étude de Popov [73], V. Iakoubovitch [95], M. Aïzerman et F. Gantmacher [3].

Pour le cas principal examiné ci-dessus le résultat définitif a été obtenu par Iakoubovitch [95] qui avait démontré le théorème suivant lequel la *condition de Popov* (36) *est dans le cas principal une condition nécessaire et suffisante d'existence des fonctions de Liapounov du type* (118).

Ce théorème implique que si la stabilité absolue du système (1) peut être établie à l'aide de la fonction de Liapounov (118), il existe un nombre positif réel q tel que la condition (36) du théorème de Popov est observée. Ainsi, le théorème de Popov couvre tout l'ensemble des systèmes du type (1) dont la stabilité absolue peut être établie à l'aide de la fonction de Liapounov du type (118).

§ 7. Action des forces extérieures sur les systèmes non linéaires

1. Réduction du problème aux équations intégrales. Considérons le système décrit par les équations différentielles non linéaires suivantes

$$\dot{x}_j = \varphi_j(x_1, x_2, \ldots, x_n, t) \quad (j = 1, \ldots, n). \tag{7.1}$$

En introduisant les matrices

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad \varphi(x_1, \ldots, x_n, t) = \begin{bmatrix} \varphi_1(x_1, \ldots, x_n, t) \\ \cdots\cdots\cdots \\ \varphi_n(x_1, \ldots, x_n, t) \end{bmatrix}, \tag{7.2}$$

on peut remplacer le système d'équations différentielles scalaires (1) par l'équation différentielle vectorielle

$$\dot{x} = \varphi(x_1, \ldots, x_n, t). \tag{7.3}$$

Supposons que les fonctions $\varphi_j(x_1, \ldots, x_n, t)$ puissent être mises sous la forme

$$\varphi_j(x_1, \ldots, x_n, t) = -[u_{j1}(t)\, x_1 + u_{j2}(t)\, x_2 + \ldots \\ \ldots + u_{jn}(t)\, x_n] + y_j(t) + \psi_j(x_1, \ldots, x_n, t) \\ (j = 1, \ldots, n), \tag{7.4}$$

c'est-à-dire sous la forme de la somme d'une certaine fonction approximante

$$-\sum_{k=1}^{n} u_{jk}(t)\, x_k + y_j(t),$$

linéaire par rapport à $x_1, \ldots, x_n$ et d'une correction non linéaire $\psi_j(x_1, \ldots, x_n, t)$. L'approximation linéaire peut être réalisée de différentes façons ; on peut, par exemple, l'assujettir à la condition

de petitesse de $|\psi_j(x_1, \ldots, x_n, t)|$ dans le domaine de variation considéré de $x_1, \ldots, x_n$, etc. Le choix convenable de l'approximation linéaire rend bien plus rapide la convergence des approximations successives (35) de la solution du système d'équations (1).

Conformément à (4), le système d'équations différentielles (1) se met sous la forme

$$\dot{x}_j + \sum_{k=1}^{n} u_{jk}(t)\, x_k = y_j(t) + \psi_j(x_1, \ldots, x_n, t), \qquad (j = 1, \ldots, n). \quad (7.5)$$

Le système d'équations différentielles scalaires (5) est équivalent à l'équation différentielle vectorielle

$$\dot{x} + u(t)\, x = y(t) + \psi(x_1, \ldots, x_n, t), \quad (7.6)$$

où par $u(t)$, $y(t)$ et $\psi(x_1, \ldots, x_n, t)$ on désigne les matrices

$$\left.\begin{aligned} u(t) &= \begin{bmatrix} u_{11}(t) & \ldots & u_{1n}(t) \\ \cdot & \cdots & \cdot \\ u_{n1}(t) & \ldots & u_{nn}(t) \end{bmatrix}, \quad y(t) = \begin{bmatrix} y_1(t) \\ \cdots \\ y_n(t) \end{bmatrix}, \\ \psi(x_1, \ldots, x_n, t) &= \begin{bmatrix} \psi_1(x_1, \ldots, x_n, t) \\ \cdots \\ \psi_n(x_1, \ldots, x_n, t) \end{bmatrix}. \end{aligned}\right\} \quad (7.7)$$

Introduisons maintenant la notation $\theta(t)$ pour la matrice

$$\theta(t) = \begin{bmatrix} \theta_{11}(t) & \ldots & \theta_{1n}(t) \\ \cdot & \cdots & \cdot \\ \theta_{n1}(t) & \ldots & \theta_{nn}(t) \end{bmatrix}, \quad (7.8)$$

dont les colonnes sont des solutions particulières linéairement indépendantes de l'équation vectorielle homogène

$$\dot{x} + u(t)\, x = 0. \quad (7.9)$$

La matrice $\theta(t)$ est donc la matrice fondamentale des solutions de l'équation différentielle (9). Chaque colonne de $\theta(t)$ présentant une solution particulière de (9), la matrice $\theta(t)$ vérifie l'équation

$$\dot{\theta}(t) + u(t)\, \theta(t) = 0. \quad (7.10)$$

Cherchons la solution de l'équation différentielle vectorielle (6) sous la forme

$$x = \theta(t)\, \chi(t), \quad (7.11)$$

où

$$\chi(t) = \begin{bmatrix} \chi_1(t) \\ \cdots \\ \chi_n(t) \end{bmatrix}. \quad (7.12)$$

Le vecteur à définir est $\chi(t)$. En portant l'expression (11) dans l'équation (6) on obtient

$$\dot{\theta}(t)\,\chi(t)+\theta(t)\,\dot{\chi}(t)+u(t)\,\theta(t)\,\chi(t)=$$
$$=y(t)+\psi(x_1,\ \ldots,\ x_n,\ t). \tag{7.13}$$

En vertu de (10), le premier et le troisième terme du premier membre de l'équation (13) se réduisent et l'équation (13) devient:

$$\theta(t)\,\dot{\chi}(t)=y(t)+\psi(x_1,\ \ldots,\ x_n,\ t), \tag{7.14}$$

on en tire

$$\dot{\chi}(t)=\theta^{-1}(t)\,y(t)+\theta^{-1}(t)\,\psi(x_1,\ \ldots,\ x_n,\ t), \tag{7.15}$$

où par $\theta^{-1}(t)$ on désigne la matrice inverse. Il ressort de l'équation (15) que

$$\chi(t)=\int_{t_0}^{t}\theta^{-1}(\tau)\,y(\tau)\,d\tau+\int_{t_0}^{t}\theta^{-1}(\tau)\,\psi[x_1(\tau),\ \ldots,\ x_n(\tau),\ \tau]\,d\tau+C, \tag{7.16}$$

avec C, le vecteur des constantes arbitraires

$$C=\begin{bmatrix}C_1\\ \cdots\\ C_n\end{bmatrix}. \tag{7.17}$$

En portant dans (11) l'expression (16) du vecteur $\chi(t)$, il vient

$$x(t)=\theta(t)\,C+\theta(t)\int_{t_0}^{t}\theta^{-1}(\tau)\,y(\tau)\,d\tau+$$
$$+\theta(t)\int_{t_0}^{t}\theta^{-1}(\tau)\,\psi[x_1(\tau),\ \ldots,\ x_n(\tau),\ \tau]\,d\tau. \tag{7.18}$$

Déterminons le vecteur C à partir des conditions initiales

$$[x(t)]_{t=t_0}=x(t_0). \tag{7.19}$$

La relation (18) amène

$$x(t_0)=\theta(t_0)\,C, \tag{7.20}$$

et donc

$$C=\theta^{-1}(t_0)\,x(t_0). \tag{7.21}$$

Finalement la relation (18) se met sous la forme

$$x(t) = \theta(t)\,\theta^{-1}(t_0)\,x(t_0) + \int_{t_0}^{t} \theta(t)\,\theta^{-1}(\tau)\,y(\tau)\,d\tau + \\ + \int_{t_0}^{t} \theta(t)\,\theta^{-1}(\tau)\,\psi[x_1(\tau), \ldots, x_n(\tau), \tau]\,d\tau. \quad (7.22)$$

La relation (22) dont la fonction sous le signe somme du second membre contient les éléments du vecteur $x(t)$ est une équation intégrale vectorielle non linéaire par rapport à l'inconnue $x(t)$. Cette équation intégrale est équivalente à l'équation différentielle vectorielle (3) considérée avec ses conditions initiales (19).

Introduisons maintenant les notations

$$N(t, \tau) = \theta(t)\,\theta^{-1}(\tau). \quad (7.23)$$

La fonction $N(t, \tau)$ est une matrice carrée

$$N(t, \tau) = \begin{bmatrix} N_{11}(t, \tau), & \ldots, & N_{1n}(t, \tau) \\ \cdot\cdot\cdot & \cdot\cdot\cdot & \cdot\cdot\cdot \\ N_{n1}(t, \tau), & \ldots, & N_{nn}(t, \tau) \end{bmatrix},$$

dont les éléments sont

$$N_{jk}(t, \tau) = \sum_{\mu=1}^{n} \theta_{j\mu}(t)\,[\theta^{-1}(\tau)]_{\mu k}.$$

Notons que la fonction $N(t, \tau)$ vérifie la relation suivante:

$$N(t, \sigma) = N(t, \tau)\,N(\tau, \sigma), \quad (7.24)$$

qui a lieu quels que soient t, τ et σ. En effet, d'après (23)

$$N(t, \tau)\,N(\tau, \sigma) = \theta(t)\,\theta^{-1}(\tau)\,\theta(\tau)\,\theta^{-1}(\sigma) = \theta(t)\,\theta^{-1}(\sigma),$$

ce qui coïncide avec la relation (24).

La matrice inverse de la matrice $N(t, \tau)$ est

$$[N(t, \tau)]^{-1} = N(\tau, t). \quad (7.25)$$

La relation (25) résulte du fait que

$$N(t, \tau)\,N(\tau, t) = \theta(t)\,\theta^{-1}(\tau)\,\theta(\tau)\,\theta^{-1}(t) = E,$$

où par E on désigne la matrice unité.

La relation (23) implique également

$$N(t, t) = E$$

pour toute valeur de t.

L'équation intégrale vectorielle non linéaire (22) peut également s'écrire:

$$x(t)=N(t, t_0)\,x(t_0)+\int_{t_0}^{t} N(t, \tau)\,y(\tau)\,d\tau+$$
$$+\int_{t_0}^{t} N(t, \tau)\,\psi[x_1(\tau), \ldots, x_n(\tau), \tau]\,d\tau. \quad (7.26)$$

L'équation (26) est équivalente au système suivant d'équations intégrales scalaires non linéaires

$$x_j(t)=\sum_{k=1}^{n} N_{jk}(t, t_0)\,x_k(t_0)+\sum_{k=1}^{n}\int_{t_0}^{t} N_{jk}(t, \tau)\,y_k(\tau)\,d\tau+$$
$$+\sum_{k=1}^{n}\int_{t_0}^{t} N_{jk}(t, \tau)\,\psi_k[x_1(\tau), \ldots, x_n(\tau), \tau]\,d\tau$$
$$(j=1, \ldots, n). \quad (7.27)$$

2. **Construction des solutions approchées.** Dans le cas général, le système initial des équations différentielles non linéaires (1) n'est pas intégrable. Dans ce sens, la réduction du problème au système d'équations intégrales non linéaires (27) ne change rien au fond, du fait que dans le cas général les équations (27) ne peuvent non plus être résolues avec précision. Toutefois, pour construire les solutions approchées du système (1), il est plus commode de partir du système d'équations intégrales (27). Pour abréger l'écriture, introduisons les notations suivantes

$$g(t)=N(t, t_0)\,x(t_0)+\int_{t_0}^{t} N(t, \tau)\,y(\tau)\,d\tau, \quad (7.28)$$

$$K[t, \tau, x_1(\tau), \ldots, x_n(\tau)] = N(t, \tau)\,\psi[x_1(\tau), \ldots, x_n(\tau), \tau]. \quad (7.29)$$

La matrice fondamentale $\theta(t)$ de l'équation différentielle linéaire (9) étant supposée connue, on connaît également la fonction $N(t, \tau)$ déterminée par la formule (23). On peut donc considérer comme connue la fonction $g(t)$.

Les fonctions $g(t)$ et $K[t, \tau, x_1(\tau), \ldots, x_n(\tau)]$ sont des vecteurs. Les éléments de ces vecteurs sont

$$g_j(t) = \sum_{k=1}^{n} N_{jk}(t, t_0)\, x_k(t_0) + \sum_{k=1}^{n} \int_{t_0}^{t} N_{jk}(t, \tau)\, y_k(\tau)\, d\tau \quad (j = 1, \ldots, n), \tag{7.30}$$

$$K_j(t, \tau, x_1(\tau), \ldots, x_n(\tau)) = \sum_{k=1}^{n} N_{jk}(t, \tau)\, \psi_k(x_1(\tau), \ldots, x_n(\tau), \tau) \quad (j = 1, \ldots, n). \tag{7.31}$$

L'équation intégrale vectorielle (26) peut maintenant s'écrire

$$x(t) = g(t) + \int_{t_0}^{t} K[t, \tau, x_1(\tau), \ldots, x_n(\tau)]\, d\tau. \tag{7.32}$$

Le système d'équations intégrales scalaires qui correspond à l'équation vectorielle (32) est

$$x_j(t) = g_j(t) + \int_{t_0}^{t} K_j[t, \tau, x_1(\tau), \ldots, x_n(\tau)]\, d\tau \quad (j = 1, \ldots, n). \tag{7.33}$$

Pour résoudre le système d'équations intégrales non linéaires (33) sous les hypothèses suffisamment larges relatives à la forme des fonctions non linéaires $\psi_j(x_1, \ldots, x_n, t)$, on peut appliquer la méthode des approximations successives. Adoptons comme approximation nulle du système d'équations (33)

$$x_j^{(0)}(t) = g_j(t) \quad (j = 1, \ldots, n). \tag{7.34}$$

Les approximations successives sont

$$\left.\begin{aligned}
x_j^{(1)}(t) &= g_j(t) + \int_{t_0}^{t} K_j[t, \tau, x_1^{(0)}(\tau), \ldots, x_n^{(0)}(\tau)]\, d\tau, \\
x_j^{(2)}(t) &= g_j(t) + \int_{t_0}^{t} K_j[t, \tau, x_1^{(1)}(t), \ldots, x_n^{(1)}(\tau)]\, d\tau, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
x_j^{(m)}(t) &= g_j(t) + \int_{t_0}^{t} K_j[t, \tau, x_1^{(m-1)}(\tau), \ldots, x_n^{(m-1)}(\tau)]\, d\tau, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\qquad (j = 1, \ldots, n).
\end{aligned}\right\} \tag{7.35}$$

La méthode des approximations successives présente cet inconvénient que les quadratures des expressions (35) se compliquent rapidement à mesure que l'ordre des approximations augmente.

Une autre méthode de résolution approchée du système d'équations intégrales (33) consiste à diviser l'intervalle (t_0, t) de la façon suivante

$$\left.\begin{array}{l} \Delta t_0 = t_1 - t_0, \\ \Delta t_1 = t_2 - t_1, \\ \dots\dots\dots \\ \Delta t_k = t_{k+1} - t_k, \\ \dots\dots\dots \end{array}\right\} \quad (7.36)$$

Adoptons par approximation que dans chacun de ces intervalles les fonctions $x_j(t)$ gardent invariables celles des valeurs qu'elles possédaient à l'extrémité gauche de l'intervalle. On obtient alors à partir des équations (33)

$$\left.\begin{array}{l} x_j(t_0) = g_j(t_0), \\ x_j(t_1) = g_j(t_1) + K_j[t_1, t_0, x_1(t_0), \dots, x_n(t_0)]\,\Delta t_0, \\ x_j(t_2) = g_j(t_2) + K_j[t_2, t_0, x_1(t_0), \dots, x_n(t_0)]\,\Delta t_0 + \\ \qquad + K_j[t_2, t_1, x_1(t_1), \dots, x_n(t_1)]\,\Delta t_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_j(t_s) = g_j(t_s) + \sum\limits_{k=0}^{s-1} K_j[t_s, t_k, x_1(t_k), \dots, x_n(t_k)]\,\Delta t_k, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \qquad (j = 1, \dots, n). \end{array}\right\} \quad (7.37)$$

La méthode décrite permet de construire la solution approchée du système d'équations intégrales non linéaires (33), c'est-à-dire de trouver une loi approchée du mouvement du système (1) dans l'intervalle (t_0, t) qui nous intéresse.

§ 8. Méthodes qualitatives d'étude du mouvement des systèmes non linéaires

1. Systèmes non linéaires à un degré de liberté. Les méthodes qualitatives d'étude du mouvement des systèmes non linéaires s'appuient sur les résultats fournis par la théorie qualitative des équations différentielles [68] qui remonte aux travaux de A. Poincaré [75]. Ces méthodes ont été largement appliquées à la théorie des oscillations non linéaires entre autres par L. Mandelstam, A. Andronov, N. Krylov, N. Bogolioubov, B. Boulgakov, I. Mitropolski [4, 17, 47, 64]. Les méthodes qualitatives sont efficaces surtout dans l'étude du mouvement des systèmes non linéaires à un degré de liberté; plusieurs de ces méthodes peuvent être appliquées avec succès à l'étude des systèmes non linéaires à plusieurs degrés de liberté. Dans ce qui suit nous nous bornons à l'examen des systèmes à un degré de liberté.

L'énergie cinétique d'un système à un degré de liberté, dont les liaisons de même que la définition de la coordonnée généralisée q ne dépendent pas d'une façon explicite du temps, peut être mise sous la forme

$$T = \frac{1}{2} m(q)\dot{q}^2. \tag{8.1}$$

La fonction $m(q)$ s'appelle *masse réduite* du système. L'énergie potentielle du système est $V = V(q)$. Pour le système considéré, la fonction de Lagrange s'écrit

$$L = T - V. \tag{8.2}$$

L'équation du mouvement du système est de la forme:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = Q^*(q, \dot{q}, t) \tag{8.3}$$

ou

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}}\right) - \left(\frac{\partial T}{\partial q}\right) = -\frac{\partial V}{\partial q} + Q^*(q, \dot{q}, t),$$

avec $Q^*(q, \dot{q}, t)$ la force supplémentaire *non conservative* appliquée au système.

Dans certains problèmes il est plus commode d'utiliser des variables canoniques. Pour un système à un degré de liberté ces variables sont la coordonnée généralisée q et l'impulsion canonique (ou généralisée) p déterminée par la formule

$$p = \frac{\partial L}{\partial \dot{q}}. \tag{8.4}$$

Pour les systèmes aux liaisons indépendantes du temps, la fonction de Hamilton s'écrit

$$H = T + V. \tag{8.5}$$

Puisque d'après (1) et (2)

$$p = m(q)\dot{q}, \tag{8.6}$$

la fonction H peut s'écrire

$$H = \frac{p^2}{2m(q)} + V(q). \tag{8.7}$$

Les équations canoniques du mouvement du système à un degré de liberté sont

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{\partial H}{\partial p}, \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial q} + Q(q, p, t), \end{aligned}\right\} \tag{8.8}$$

où $Q(q, p, t)$ est une force supplémentaire non conservative ramenée aux variables canoniques.

D'après (7) les équations de mouvement canoniques (8) deviennent

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{p}{m(q)}, \\ \frac{dp}{dt} &= \frac{m'(q)}{2[m(q)]^2} p^2 - V'(q) + Q(q, p, t). \end{aligned}\right\} \quad (8.9)$$

Lorsque la force supplémentaire non conservative Q ne dépend pas explicitement du temps, le mouvement d'un système aux liaisons indépendantes du temps porte le nom *d'oscillations libres* du système. Dans ce cas les équations canoniques (9) s'écrivent

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{p}{m(q)}, \\ \frac{dp}{dt} &= \frac{m'(q)}{2[m(q)]^2} p^2 - V'(q) + Q(q, p). \end{aligned}\right\} \quad (8.10)$$

Les valeurs instantanées de q et p déterminent la position et la vitesse du système à un degré de liberté ou son état (phase). q et p peuvent être interprétées comme des coordonnées cartésiennes d'un point (point représentatif) dans le plan qp. Ce dernier s'appelle *plan de phase*; q et p sont les coordonnées de phase du point représentatif dans le plan de phase. De plus $\frac{dq}{dt}$ et $\frac{dp}{dt}$ sont les composantes de la vitesse du point représentatif. Cette vitesse est dite *vitesse de phase*.

Dans le plan qp, toute solution de l'équation (10)

$$q = q(t), \qquad p = p(t)$$

détermine une certaine trajectoire suivant laquelle se déplace le point représentatif. Cette trajectoire s'appelle *trajectoire de phase*. (Il ne faut pas confondre les trajectoires de phase et les vitesses de phase avec les trajectoires et les vitesses des points réels du système matériel considéré.) L'équation différentielle qui détermine les trajectoires de phase peut s'obtenir sous une forme explicite en éliminant la variable t des équations canoniques (10). Cette équation s'écrit:

$$\frac{dp}{dq} = \frac{\dfrac{m'(q)}{2[m(q)]^2} p^2 - V'(q) + Q(q, p)}{\dfrac{p}{m(q)}}. \quad (8.11)$$

L'intégration de (11) donne une équation de la famille des courbes intégrales

$$p = p(q; C) \quad (8.12)$$

qui dépend du paramètre C.

Si l'équation (11) est intégrée, le temps pendant lequel le point représentatif parcourt la trajectoire de phase est déterminé par une quadrature. En effet, la première équation (10) et l'équation (12) entraînent

$$dt = \frac{m(q)\,dq}{p(q;\ C)}, \tag{8.13}$$

d'où l'on tire

$$t = \int \frac{m(q)\,dq}{p(q;\ C)} + C_1, \tag{8.14}$$

où C_1 est une constante arbitraire. Pour déterminer les constantes arbitraires C et C_1 il faut se donner les conditions initiales, c'est-à-dire les valeurs q et p à l'instant initial $t = t_0$. C'est ce qui détermine la loi du mouvement du système.

Si avec le temps t qui varie de $-\infty$ à ∞ le point représentatif parcourt toute la courbe intégrale (12) qui correspond à une valeur fixe du paramètre C, cette courbe intégrale est une trajectoire de phase. Des cas sont pourtant possibles (décrits dans ce qui suit) où la courbe intégrale est composée de plusieurs trajectoires de phase, c'est-à-dire t variant de $-\infty$ à ∞, le point représentatif ne parcourt qu'une seule branche de la courbe intégrale (12).

Les points (q^*, p^*) du plan de phase en lesquels les deux composantes de la vitesse de phase $\frac{dq}{dt}$ et $\frac{dp}{dt}$ s'annulent simultanément, s'appellent *points singuliers*. Ce sont des points où l'équation (11) ne définit pas la pente de la trajectoire de phase par rapport à l'axe des abscisses, du fait que d'après (11) aux points (q^*, p^*) on a $\frac{dp}{dq} = \frac{0}{0}$.

Les valeurs (q^*, p^*) des coordonnées de phase des points singuliers s'obtiennent à partir des équations finies

$$\left.\begin{aligned} &\frac{p}{m(q)} = 0, \\ &\frac{m'(q)}{2\,[m(q)]^2}\,p^2 - V'(q) + Q(q,\ p) = 0. \end{aligned}\right\} \tag{8.15}$$

Puisque la fonction $m(q)$ ne peut prendre que des valeurs finies positives (la fonction $m(q)$ est la masse réduite du système), la première équation (15) donne

$$p^* = 0, \tag{8.16}$$

c'est-à-dire les points singuliers reposent sur l'axe des abscisses du plan de phase. La deuxième équation (15) et la relation (16) entraînent que les abscisses q^* des points singuliers sont les racines de l'équation

$$-V'(q) + Q(q,\ 0) = 0. \tag{8.17}$$

L'état du système

$$q = q^* = \text{const}, \qquad p = p^* = 0 \tag{8.18}$$

constitue une *solution particulière* des équations canoniques du mouvement (10) du système. Cette solution particulière correspond à la position d'équilibre du système considéré. Ainsi, les points singuliers déterminent la *position d'équilibre* du système.

A l'exception des points singuliers, la vitesse de phase diffère du zéro quel que soit le point du plan de phase, c'est-à-dire le point représentatif parcourt la trajectoire de phase sans marquer d'arrêt. Aux points d'intersection de la trajectoire avec l'axe des abscisses (points réguliers) on a d'après (11)

$$\left(\frac{dq}{dp}\right)_{p=0} = 0, \quad \left(\frac{d^2q}{dp^2}\right)_{p=0} \neq 0. \tag{8.19}$$

Ainsi la trajectoire de phase coupe l'axe des abscisses en un point régulier (c'est-à-dire non singulier) sous un angle droit et se situe d'un côté par rapport à la tangente verticale.

Comme $m(q) > 0$, dans le demi-plan supérieur du plan de phase $\dot{q} > 0$; la trajectoire de phase est suivie dans le sens des valeurs croissantes de q. Dans le demi-plan inférieur $\dot{q} < 0$ et la trajectoire est parcourue dans le sens des valeurs décroissantes de q.

2. Systèmes conservatifs. Si les forces appliquées à un système possèdent un potentiel, c'est-à-dire si elles sont déterminées par l'expression $-V'(q)$, et si ce système ne subit pas l'action d'autres forces

$$Q(q, p, t) \equiv 0, \tag{8.20}$$

un tel système s'appelle *conservatif*.

Conformément à (8) les équations canoniques du mouvement d'un système conservatif à un degré de liberté sont

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{\partial H}{\partial p}, \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial q}. \end{aligned}\right\} \tag{8.21}$$

En tenant compte de l'expression (7) de la fonction de Hamilton H, ramenons l'équation (21) à la forme

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{p}{m(q)}, \\ \frac{dp}{dt} &= \frac{m'(q)}{2[m(q)]^2} p^2 - V'(q). \end{aligned}\right\} \tag{8.22}$$

Comme d'après (21)

$$\frac{dH}{dt} = \frac{\partial H}{\partial q}\frac{dq}{dt} + \frac{\partial H}{\partial p}\frac{dp}{dt} = 0,$$

on a la première intégrale (intégrale d'énergie)

$$H = \text{const}, \tag{8.23}$$

qui d'après (7) peut être mise sous la forme

$$\frac{p^2}{2m(q)} + V(q) = v. \tag{8.24}$$

Ici v est la valeur de l'énergie totale du système, qui peut être déterminée à partir des conditions initiales: $q = q(t_0)$, $p = p(t_0)$ avec $t = t_0$.

La relation (24) peut être considérée comme une équation finie des trajectoires de phase dépendant du paramètre v. Comme il suit de (24)

$$p = \pm \sqrt{2m(q)[v - V(q)]}, \tag{8.25}$$

c'est-à-dire les trajectoires de phase d'un système conservatif sont symétriques par rapport à l'axe des abscisses.

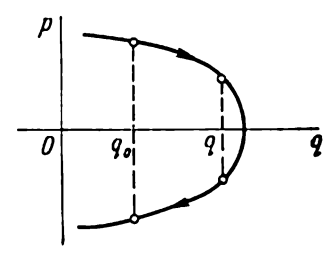


Fig. 8.1

En vertu de (14) et (25), le temps nécessaire à un point représentatif pour se déplacer du point d'abscisse q_0 au point d'abscisse q de la trajectoire de phase (fig. 8.1) est

$$t = \pm \int_{q_0}^{q} \frac{\sqrt{m(q)}\, dq}{\sqrt{2[v - V(q)]}}. \tag{8.26}$$

Dans l'expression (26) le signe supérieur devant l'intégrale correspond au cas où le point représentatif se déplace suivant l'arc situé dans le demi-plan supérieur, et le signe inférieur, suivant l'arc situé dans le demi-plan inférieur.

Un exemple des trajectoires de phase d'un système conservatif est donné par la figure 8.2. En partant de la courbe donnée de la fonction $V(q)$ (fig. 8.2, *a*), on a construit ici les courbes des fonctions $p^2 = 2m(q)[v_i - V(q)]$ (fig. 8.2, *b*) aux différentes valeurs v_i ($i = 1, \ldots, 5$) de l'énergie totale du système. Ces courbes ont

permis de construire sur la figure 8.2, *c* les trajectoires de phase associées aux valeurs v_i ($i = 1, \ldots, 5$) de l'énergie totale du système, données par la figure 8.2, *a*.

Considérons l'allure des mouvements du système qui correspondent aux différentes trajectoires de phase visualisées par la figure 8.2, *c*. Aux points *A*, *B*, *C*, où la fonction $V(q)$ a son extrémum, correspondent des points singuliers repérés respectivement sur la figure 8.2, *c* par les chiffres *2*, *4* et *1*. Comme nous avons dit plus haut, les points singuliers déterminent la position d'équilibre du système.

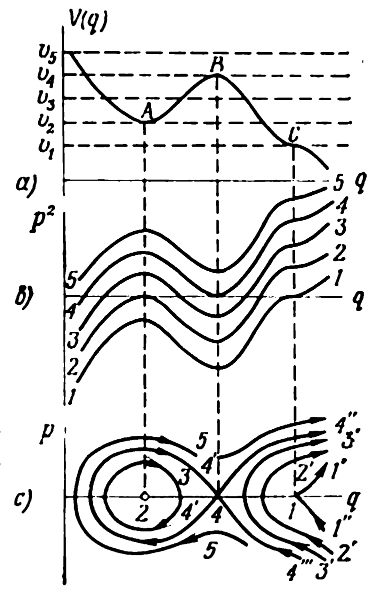


Fig. 8.2

En *A* la fonction $V(q)$ possède un minimum. Le point singulier *2* qui lui correspond est un point singulier *isolé*. Ce point s'appelle *sommet*. Si à l'instant initial le point représentatif est assez proche du point singulier *2*, il parcourt, comme le montre la figure 8.2, *c*, une trajectoire de phase fermée reposant dans un voisinage assez proche du point *2*. Un continuum de trajectoires de phase fermées associées au paramètre v_i et analogues à *3* couvre le domaine tout entier du plan de phase délimité par la courbe fermée *4'*. Ainsi, la position d'équilibre qui correspond à un point singulier du type sommet est *stable*, ce qui vérifie le théorème connu de Lejeune-Dirichlet.

En *B* la fonction $V(q)$ possède un maximum. Le point singulier *4* qui lui correspond est celui par lequel passe la courbe *4'''4'4''* qui se coupe et qui s'appelle *séparatrice*. Le point singulier *4* se nomme *col*. La séparatrice se compose de trois branches *4'''*, *4'* et *4''*. Comme nous allons le montrer plus loin, en parcourant une séparatrice quand $t \to \infty$, le point représentatif s'approche asymptotiquement du point singulier *4*. C'est pourquoi chacune des branches isolées de la séparatrice (*4'''*, *4'* et *4''*) susmentionnées est une trajectoire de phase. De plus, la trajectoire de phase *4'''* est parcourue en un mouvement *infini limité* (pour $t \to \infty$, $q \to q_4$; pour $t \to -\infty$, $q \to \infty$). La trajectoire de phase *4''* est parcourue en un mouvement *limité infini* (pour $t \to \infty$, $q \to \infty$; pour $t \to -\infty$, $q \to q_4$). La trajectoire de phase *4'* est parcourue en un mouvement *limité double* (pour $t \to \infty$, $q \to q_4$; pour $t \to -\infty$, $q \to q_4$).

Si à l'instant initial le point représentatif se trouve aussi près que l'on veut du point singulier *4*, il suivra alors, comme on le

voit sur la figure 8.2, *c*, l'une des trajectoires de phase du type *3*, *5* ou *3'*, mais comme pour le mouvement suivant les trajectoires de phase du type *5* et *3'* $q \to \infty$ lorsque $t \to \infty$, la position d'équilibre associée au point singulier *4* (c'est-à-dire au point singulier du type col) est *instable*.

En C la fonction $V(q)$ a une valeur stationnaire. Le point singulier *1* qui lui correspond est (comme nous le montrons dans ce qui suit) un point de rebroussement de première espèce. Ainsi qu'il ressort de la disposition des trajectoires de phase de la figure 8.2, *c*, la position d'équilibre relative au point singulier *1* est également instable.

La séparatrice qui passe par le point singulier *1* est composée de deux branches *1'* et *1"* dont chacune est une trajectoire de phase du système. La trajectoire de phase *1"* est parcourue en un mouvement infini limité (pour $t \to \infty$, $q \to q_1$; pour $t \to -\infty$, $q \to \infty$). La trajectoire de phase *1'* est suivie par un mouvement limité infini (pour $t \to \infty$, $q \to \infty$; pour $t \to -\infty$, $q \to q_1$).

Les trajectoires de phase du type *5*, *3'*, *2'* correspondent aux mouvements infinis doubles (pour $t \to \infty$, $q \to \infty$; pour $t \to -\infty$, $q \to \infty$).

La trajectoire de phase *3* est une courbe fermée. En tout point de cette trajectoire la vitesse de phase est différente du zéro; le point représentatif suit la trajectoire *3* sans s'arrêter et effectue un tour en un intervalle de temps fini. La trajectoire *3* correspond à un mouvement de *libration* (pulsé) du système.

Déterminons maintenant la pente d'une trajectoire de phase par rapport à l'axe des abscisses en un point singulier. L'expression (25) conduit à

$$\frac{dp}{dq} = \pm \frac{m'(q)\,[v - V(q)] - m(q)\,V'(q)}{\sqrt{2m(q)\,[v - V(q)]}}. \tag{8.27}$$

Nous avons montré ci-dessus (16) qu'en un point singulier $p = 0$. D'après (17), dans un système conservatif, $V'(q)$ s'annule en un point singulier. Ainsi, en tenant compte de l'expression (25), on a les relations:

$$v = V(q^*), \qquad V'(q^*) = 0, \tag{8.28}$$

où q^* est l'abscisse du point singulier.

Soit μ l'ordre de la plus petite des dérivées de $V(q)$, différente de zéro au point singulier

$$V^{(\mu)}(q^*) \neq 0.$$

On peut alors mettre les fonctions $V(q)$ et $V'(q)$ sous la forme

$$\left.\begin{aligned} V(q) &= V(q^*) + \frac{1}{\mu!}\,[V^{(\mu)}(q^*) + \varepsilon(q)]\,(q - q^*)^{\mu}, \\ V'(q) &= \frac{1}{(\mu-1)!}\,[V^{(\mu)}(q^*) + \eta(q)]\,(q - q^*)^{\mu-1}, \end{aligned}\right\} \tag{8.29}$$

où les fonctions $\varepsilon(q)$ et $\eta(q)$ s'annulent en $q = q^*$:

$$\varepsilon(q^*) = 0, \qquad \eta(q^*) = 0. \tag{8.30}$$

Conformément à (29), l'expression (27) peut être récrite

$$\frac{dp}{dq} = \pm \frac{\mathcal{R}}{\sqrt{-2m(q)\mathcal{F}}\,(q-q^*)^{\frac{\mu}{2}}}, \tag{8.31}$$

avec

$$\left.\begin{array}{l} \mathcal{F} = \frac{1}{\mu!}[V^{(\mu)}(q^*) + \varepsilon(q)], \quad \mathcal{Q} = \frac{1}{(\mu-1)!}[V^{(\mu)}(q^*) + \eta(q)], \\ \mathcal{R} = -m'(q)\mathcal{F}(q-q^*)^{\mu} - m(q)\mathcal{Q}(q-q^*)^{\mu-1}. \end{array}\right\} \tag{8.32}$$

Il résulte de la relation (28) que $\mu \geqslant 2$. Lorsque $\mu = 2$ et $V''(q^*) > 0$, la fonction $V(q)$ possède en $q = q^*$ un minimum, et par suite, le point $p = 0$, $q = q^*$ est un point singulier isolé.

Dans le cas où $\mu = 2$ et $V''(q^*) < 0$, la fonction $V(q)$ possède un maximum en $q = q^*$. Le point $p = 0$, $q = q^*$ est un point singulier du type *col*. On a alors d'après (31)

$$\lim_{q \to q^*} \frac{dp}{dq} = \pm k, \qquad k = \sqrt{-m(q^*)V''(q^*)}. \tag{8.33}$$

Comme dans le cas considéré $V''(q^*) < 0$, l'expression sous le signe de la racine de (33) est positive et k est une valeur réelle.

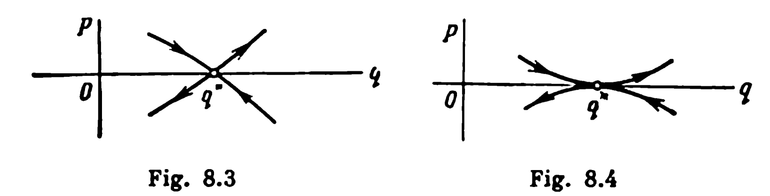


Fig. 8.3 Fig. 8.4

Ainsi, nous sommes ici dans le cas où la séparatrice qui passe par le point singulier se coupe en ce point (fig. 8.3). Pour $\mu > 2$ pair ($\mu = 4, 6, 8, \ldots,$) et $V^{(\mu)}(q^*) < 0$, la fonction $V(q)$ possède en $q = q^*$ un maximum. Dans ce cas on a, d'après (31),

$$\lim_{q \to q^*} \frac{dp}{dq} = 0, \tag{8.34}$$

et, par conséquent, une *auto-ausculation* de la séparatrice qui passe par le point singulier (fig. 8.4). Si $\mu > 2$ et est un nombre impair ($\mu = 3, 5, 7, \ldots,$), en $q = q^*$ la fonction $V(q)$ possède un minimax. Dans ce cas, en vertu de (31),

$$\lim_{q \to q^*} \frac{dp}{dq} = 0, \tag{8.35}$$

et le point singulier est un point de rebroussement de première espèce. La séparatrice qui passe par le point singulier est représentée sur la figure 8.5.

Déterminons maintenant le temps nécessaire au point représentatif pour se déplacer suivant la séparatrice du point d'abscisse $q = q_0$ au point singulier d'abscisse $q = q^*$ (fig. 8.6). Supposons qu'à l'instant $t = t_0$ le point représentatif se trouve sur la branche

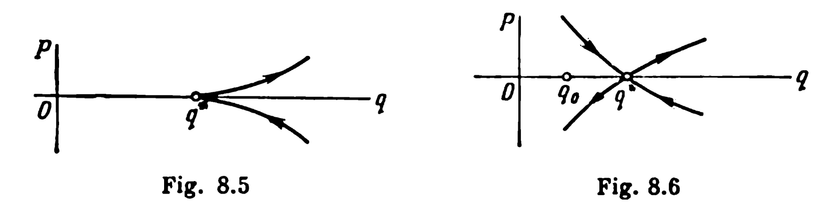


Fig. 8.5 Fig. 8.6

de la séparatrice dans le demi-plan supérieur au point d'abscisse q_0. D'après (26) le temps nécessaire au point représentatif pour atteindre le point singulier est

$$t^* = \int_{q_0}^{q^*} \frac{\sqrt{m(q)}\,dq}{\sqrt{2[v - V(q)]}}. \tag{8.36}$$

En portant dans (36) l'expression (29) qui détermine la fonction $V(q)$, on obtient

$$t^* = \int_{q_0}^{q^*} \frac{\sqrt{m(q)}}{\sqrt{-\frac{2}{\mu!}[V^{(\mu)}(q^*) + \varepsilon(q)]}} \frac{dq}{(q - q^*)^{\frac{\mu}{2}}}. \tag{8.37}$$

Puisque $\frac{\mu}{2} \geqslant 1$, l'intégrale impropre (37) est divergente, et, par suite, comme nous l'avons dit plus haut, avec $t \to \infty$, le point représentatif s'approche asymptotiquement du point singulier.

3. Systèmes dissipatifs. On appelle dissipatifs les systèmes soumis, en plus des forces conservatives, à des forces *dissipatives*, c'est-à-dire à des forces dont l'action disperse l'énergie du système.

Considérons un système à un degré de liberté dans lequel

$$m(q) = m = \text{const}, \tag{8.38}$$

$$V(q) = \frac{1}{2} cq^2, \tag{8.39}$$

$$Q(q, p) = -cB \operatorname{sign} p - 2\varepsilon p. \tag{8.40}$$

La force supplémentaire $Q(q, p)$ est la somme des forces de frottement sec et de frottement visqueux.

D'après (10), les équations canoniques du mouvement du système sont

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{p}{m}, \\ \frac{dp}{dt} &= -cq - cB\operatorname{sign} p - 2\varepsilon p. \end{aligned}\right\} \tag{8.41}$$

Le système d'équations différentielles (41) peut être remplacé par une équation différentielle du second ordre

$$m\ddot{q} = -cq - cB\operatorname{sign}\dot{q} - 2\varepsilon m\dot{q}. \tag{8.42}$$

En désignant

$$k^2 = \frac{c}{m}, \tag{8.43}$$

mettons l'équation (42) sous la forme

$$\frac{d^2}{dt^2}(q+B) + 2\varepsilon\frac{d}{dt}(q+B) + k^2(q+B) = 0 \quad (\text{pour } \dot{q} > 0), \tag{8.44}$$

$$\frac{d^2}{dt^2}(q-B) + 2\varepsilon\frac{d}{dt}(q-B) + k^2(q-B) = 0 \quad (\text{pour } \dot{q} < 0). \tag{8.44'}$$

Etudions le mouvement du système sous les conditions initiales suivantes

$$q = a_j > 0, \qquad \dot{q} = 0 \qquad \text{pour} \qquad t = 0. \tag{8.45}$$

Pour passer à la définition de la loi qui régit le mouvement du système, supposons qu'à l'instant qui suit l'instant initial le système se déplace avec $\dot{q} < 0$, c'est-à-dire que le point représentatif se déplace dans le demi-plan inférieur du plan de phase. On a alors l'équation (44')

$$\frac{d^2}{dt^2}(q-B) + 2\varepsilon\frac{d}{dt}(q-B) + k^2(q-B) = 0.$$

Une fois cette équation intégrée, il faut justifier l'hypothèse d'après laquelle $\dot{q} < 0$ et déterminer l'instant où le point représentatif passe dans le demi-plan supérieur.

Sous la condition $\varepsilon < k$, les racines de l'équation caractéristique

$$\lambda^2 + 2\varepsilon\lambda + k^2 = 0 \tag{8.46}$$

sont

$$\lambda_1, \ \lambda_2 = -\varepsilon \pm i\omega, \quad \omega = \sqrt{k^2 - \varepsilon^2}. \tag{8.47}$$

Sous les conditions initiales (45), la solution de l'équation différentielle (44') s'écrit

$$q(t) = B + (a_j - B)\,e^{-\varepsilon t}\cos\omega t + \frac{\varepsilon}{\omega}(a_j - B)\,e^{-\varepsilon t}\sin\omega t. \tag{8.48}$$

On en tire

$$\dot{q}(t) = -\left(\frac{\varepsilon^2}{\omega} + \omega\right)(a_j - B)\, e^{-\varepsilon t} \sin \omega t. \tag{8.49}$$

L'expression (49) montre que $\dot{q} < 0$ dans l'intervalle de temps $0 < t < t_1$, où

$$t_1 = \frac{\pi}{\omega}, \tag{8.50}$$

ce qui justifie par conséquent l'hypothèse que le mouvement du point représentatif à l'instant qui suit l'instant initial a lieu dans le demi-plan de phase inférieur.

Ainsi, l'expression (48) détermine la loi du mouvement du système dans l'intervalle de temps $0 < t < \frac{\pi}{\omega}$. A l'instant $t_1 = \frac{\pi}{\omega}$, la vitesse généralisée $\dot{q}(t_1) = 0$ et le point représentatif coupe l'axe des abscisses pour passer dans le demi-plan de phase supérieur.

Il résulte de (48) que la valeur de la coordonnée généralisée q à l'instant t_1 est

$$q(t_1) = B - (a_j - B)\,\beta, \tag{8.51}$$

où

$$\beta = e^{-\varepsilon \frac{\pi}{\omega}} < 1. \tag{8.52}$$

En introduisant les notations

$$a_{j+1} = \beta\left[a_j - B\left(1 + \frac{1}{\beta}\right)\right], \tag{8.53}$$

on obtient d'après (51)

$$q(t_1) = -a_{j+1}.$$

Si l'écart initial $a_j > B\left(1 + \frac{1}{\beta}\right)$, il vient

$$a_{j+1} > 0.$$

Le mouvement ultérieur du point représentatif s'effectue dans le demi-plan de phase supérieur, c'est-à-dire on vérifie l'équation (44)

$$\frac{d^2}{dt^2}(q + B) + 2\varepsilon \frac{d}{dt}(q + B) + k^2 (q + B) = 0. \tag{8.54}$$

Cette équation doit être intégrée sous les conditions initiales suivantes:

$$q = -a_{j+1} < 0, \quad \dot{q} = 0 \quad \text{pour} \quad t = t_1 = \frac{\pi}{\omega}. \tag{8.55}$$

La solution de (44) sous les conditions initiales (55) s'écrit

$$q(t) = -B - \frac{\varepsilon}{\omega}(a_{j+1} - B)\, e^{-\varepsilon(t - t_1)} \sin \omega (t - t_1) - \\ - (a_{j+1} - B)\, e^{-\varepsilon(t - t_1)} \cos \omega (t - t_1). \tag{8.56}$$

L'expression (56) conduit à

$$\dot{q}(t)=\left(\frac{\varepsilon^2}{\omega}+\omega\right)(a_{j+1}-B)\,e^{-\varepsilon(t-t_1)}\sin\omega(t-t_1). \qquad (8.57)$$

Il résulte de l'expression (57) que $\dot{q}>0$ dans l'intervalle $t_1<t<t_2$, où

$$t_2=\frac{2\pi}{\omega}, \qquad (8.58)$$

et dans cet intervalle l'expression (56) définit la loi du mouvement du système.

A l'instant t_2, la vitesse généralisée $\dot{q}(t_2)=0$ et le point représentatif, en passant par l'axe des abscisses, se déplace dans le demi-plan inférieur.

D'après (56), à l'instant t_2, la valeur de la coordonnée généralisée q est

$$q(t_2)=a_{j+2}, \qquad (8.59)$$

où

$$a_{j+2}=\beta\left[a_{j+1}-B\left(1+\frac{1}{\beta}\right)\right]. \qquad (8.60)$$

L'expression (60) montre que si $a_{j+1}>B\left(1+\frac{1}{\beta}\right)$, $q(t_2)>0$.

Les grandeurs a_j, a_{j+1}, a_{j+2}, . . . sont les valeurs absolues des abscisses des points d'intersection consécutifs de la trajectoire de phase avec l'axe $p=0$. Appelons ces grandeurs amplitudes successives. La comparaison des expressions (53) et (60) révèle que la formule associant deux amplitudes successives est la même dans le cas du point représentatif passant du demi-plan supérieur dans le demi-plan inférieur ou inversement.

Si on désigne maintenant par a_0 l'amplitude initiale, on obtient d'après (53) et (60) une suite d'amplitudes déterminées par les relations

$$\left.\begin{aligned} a_1&=\beta a_0-B(1+\beta),\\ a_2&=\beta a_1-B(1+\beta),\\ a_3&=\beta a_2-B(1+\beta),\\ &\dots\dots\dots\dots\\ a_{n-1}&=\beta a_{n-2}-B(1+\beta),\\ a_n&=\beta a_{n-1}-B(1+\beta). \end{aligned}\right\} \qquad (8.61)$$

Montrons maintenant que le mouvement du système s'effectue en un intervalle de temps fini et cesse à l'instant où le point représentatif tombe sur le segment $[-B, +B]$ de l'axe des abscisses

du plan de phase (fig. 8.7). Le segment $[-B, +B]$ de l'axe des abscisses du plan de phase s'appelle *zone inefficace* du système. En effet, en un point quelconque de ce segment, $|q| \leqslant B$ et la force de rappel cq est équilibrée par le frottement sec. Comme sur l'axe des abscisses $\dot{q} = 0$ et le frottement visqueux $-2\varepsilon m\dot{q}$ s'annule, lorsque le point représentatif passe dans la zone inefficace, la force de rappel cq est compensée par le frottement sec d'équilibre, et le mouvement s'arrête. Puisque, comme il en résulte des formules (53) et (60), avec chaque demi-oscillation l'amplitude des oscillations décroît à une grandeur dépassant $2B$, quelle que soit l'amplitude initiale a_0, après un nombre fini de demi-oscillations l'amplitude du système prend une valeur plus petite que B et le mouvement s'arrête. Ainsi, le mouvement du système s'effectue en un intervalle de temps fini.

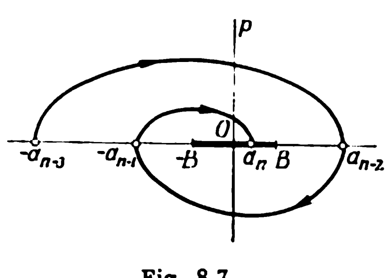


Fig. 8.7

Considérons maintenant les relations (61). Désignons par n le nombre de demi-oscillations au cours desquelles s'effectue le mouvement, pour obtenir (fig. 8.7)

$$a_n < B. \tag{8.62}$$

En multipliant la première relation (61) par β^{n-1}, la deuxième relation par β^{n-2}, etc., et en additionnant respectivement les premiers et les seconds membres des relations obtenues, il vient

$$a_n = \beta^n a_0 - B(1+\beta^n) - 2B\beta(1+\beta+\ldots+\beta^{n-2})$$

ou

$$a_n = \beta^n a_0 - B(1+\beta^n) - 2B\beta\,\frac{1-\beta^{n-1}}{1-\beta}. \tag{8.63}$$

Lorsque $\varepsilon = 0$, c'est-à-dire si les forces dissipatives appliquées au système ne sont composées que de celle du frottement sec, on a

$$\beta = 1, \quad \lim_{\beta\to 1}\frac{1-\beta^{n-1}}{1-\beta} = n-1, \tag{8.64}$$

et la formule (63) devient

$$a_n = a_0 - 2Bn. \tag{8.65}$$

Mais si le système considéré ne comporte pas de forces de frottement sec ($B = 0$), la formule (63) se met sous la forme

$$a_n = \beta^n a_0. \tag{8.66}$$

En l'absence du frottement sec, le système ne possède pas de zone inefficace et ses oscillations libres s'amortissent asymptotiquement pour $t \to \infty$.

D'après les expressions (48) et (56) la demi-période des oscillations du système vaut $\frac{\pi}{\omega}$. La période des oscillations du système s'écrit

$$P = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{k^2 - \varepsilon^2}}. \tag{8.67}$$

On voit de l'expression (67) que les forces de frottement sec n'interviennent pas dans la grandeur de la période des oscillations libres. Toutefois, nous n'avons démontré ce résultat que pour les systèmes à un degré de liberté.

4. Systèmes à auto-oscillations. Méthode des transformations ponctuelles. Passons maintenant à l'étude des systèmes soumis, en plus des forces à potentiel et des forces dissipatives, aux forces dont l'action restitue l'énergie du système. D'après Kelvin ces forces sont dites artificielles. La présence des forces artificielles suppose que le système soit lié à une source d'énergie extérieure, ces forces artificielles assurant l'utilisation de l'énergie de la source extérieure pour restituer les pertes d'énergie du système.

Si les forces dissipatives ou les forces artificielles (ou encore les unes et les autres) sont des forces non linéaires, le système peut donner lieu à des oscillations périodiques dont les paramètres ne dépendent pas des conditions initiales mais s'établissent automatiquement de façon que l'accroissement de l'énergie du système en une période d'oscillations soit nul. Les oscillations périodiques considérées (on les appelle également mouvements périodiques) sont des solutions particulières périodiques asymptotiquement stables des équations différentielles du mouvement du système. Outre ces solutions, les équations du mouvement du système peuvent avoir des solutions particulières périodiques instables. Les mouvements périodiques asymptotiquement stables s'appellent *auto-oscillations*; les systèmes non linéaires qui donnent lieu aux mouvements de ce type se nomment *systèmes à auto-oscillations*. Parmi les systèmes bien étudiés de ce type il y a les montres, les oscillateurs à tubes, etc.

Dans ce qui suit nous examinerons les propriétés caractéristiques des systèmes à auto-oscillations sur l'exemple d'un oscillateur à tubes.

Les oscillateurs à tubes sont prévus pour générer dans un circuit électrique des oscillations entretenues, c'est-à-dire du courant alternatif haute fréquence. Ils s'emploient surtout en radiotechnique. Ce type d'oscillateur est schématisé figure 8.8. La source d'énergie extérieure est une batterie d'anode E_a qui alimente un tube

électronique (triode). Le circuit oscillant RLC est intercalé dans le circuit anodique du tube. Le circuit de grille comporte une bobine de self. Le coefficient d'induction mutuelle des selfs dans le circuit de grille et le circuit anodique est noté M. Comme nous le montrons plus loin, la liaison entre le circuit anodique et le circuit de grille, assurée par l'induction mutuelle, effectue le dosage de la restitution (à partir de la batterie d'anode) de l'énergie dans le circuit RLC de façon à compenser les pertes d'énergie dans la résistance ohmique et à entretenir les oscillations.

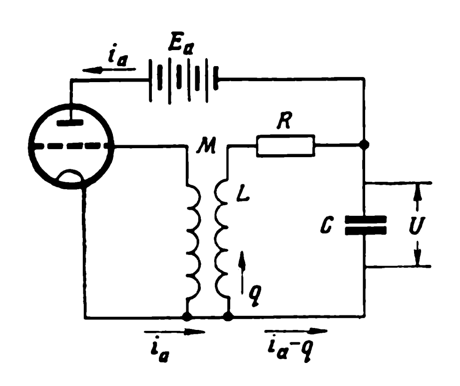


Fig. 8.8

Pour composer une équation différentielle à laquelle satisfait l'intensité du courant q dans la branche LR du circuit oscillant, déterminons la tension U sur l'armature du condensateur. En désignant par i_a le courant anodique, on a

$$U=\frac{1}{C}\int_0^t [i_a(\tau)-q(\tau)]\,d\tau. \tag{8.68}$$

Comme cette grandeur est égale à la chute de tension dans le self L et dans la résistance ohmique R, on a la relation

$$U=L\frac{dq}{dt}+Rq. \tag{8.69}$$

Les relations (68) et (69) entraînent

$$L\dot{q}+Rq=\frac{1}{C}\int_0^t [i_a(\tau)-q(\tau)]\,d\tau. \tag{8.70}$$

En dérivant par rapport à t les premier et second membres de la relation (70), on obtient l'équation différentielle suivante:

$$\ddot{q}+2\varepsilon\dot{q}+k^2q=k^2i_a, \tag{8.71}$$

où

$$\varepsilon=\frac{1}{2}\frac{R}{L},\quad k^2=\frac{1}{LC}. \tag{8.72}$$

Le courant anodique i_a est fonction de la tension de commande u:

$$i_a=f(u). \tag{8.73}$$

La courbe de la fonction $f(u)$ s'appelle caractéristique du tube. La tension de commande u est de la forme

$$u = U_g + \gamma U_a, \tag{8.74}$$

où U_g est la différence de potentiels entre la grille du tube et la cathode, et U_a, la différence de potentiels entre l'anode et la cathode. Par γ on désigne la perméabilité du tube. La courbe caractéristique

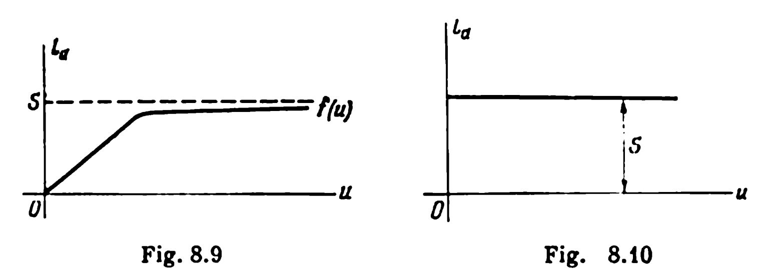


Fig. 8.9 Fig. 8.10

du tube est donnée figure 8.9. Avec l'augmentation de u le courant anodique i_a s'approche asymptotiquement du courant de saturation dont la valeur est désignée par S.

Les paramètres d'un oscillateur à tubes sont généralement tels que la relation (74) peut être approximée de la façon suivante:

$$u \approx U_g. \tag{8.75}$$

Pour simplifier les calculs, approximons encore la courbe caractéristique de l'oscillateur par une fonction en escalier visualisée figure 8.10. On peut montrer qu'en remplaçant la caractéristique du tube par une fonction en escalier on ne déforme pas l'allure des solutions de l'équation différentielle (71). Lorsque la nécessité se présente, on peut préciser les solutions obtenues en approximant la caractéristique du tube (fig. 8.9) par une ligne brisée.

L'approximation de la courbe caractéristique d'un tube par une fonction en escalier, en tenant compte de (75), donne

$$i_a = \begin{cases} S & \text{pour } U_g \geqslant 0, \\ 0 & \text{pour } U_g < 0. \end{cases} \tag{8.76}$$

Puisque l'induction mutuelle des selfs fait que dans le circuit de grille et le circuit anodique

$$U_g = M \frac{dq}{dt}, \tag{8.77}$$

la relation (76) peut s'écrire

$$i_a = \begin{cases} S & \text{pour } \dot{q} \geqslant 0, \\ 0 & \text{pour } \dot{q} < 0. \end{cases} \tag{8.78}$$

De cette façon l'équation différentielle (71) devient

$$\ddot{q}+2\varepsilon\dot{q}+k^2q=\begin{cases} k^2S & \text{pour } \dot{q}\geqslant 0, \\ 0 & \text{pour } \dot{q}<0. \end{cases} \quad (8.79)$$

Passons à l'intégration de l'équation (79), cette intégration s'effectuant par rapport aux intervalles de temps pendant lesquels la fonction $\dot{q}(t)$ garde son signe constant. Soit

$$q = a > 0, \quad \dot{q} = 0 \quad \text{pour} \quad t = 0. \quad (8.80)$$

Comme dans le mouvement à l'instant qui suit $t = 0$ $\dot{q}(t)$ est négatif (ce qui sera confirmé ci-dessous), il faut, conformément à (78), faire appel à l'équation

$$\ddot{q} + 2\varepsilon\dot{q} + k^2q = 0. \quad (8.81)$$

Sous réserve que

$$\varepsilon < k,$$

la solution de l'équation (81) qui vérifie les conditions initiales (80) est

$$q(t)=\frac{\varepsilon a}{\omega}e^{-\varepsilon t}\sin\omega t+ae^{-\varepsilon t}\cos\omega t, \quad (8.82)$$

où

$$\omega=\sqrt{k^2-\varepsilon^2}. \quad (8.83)$$

On tire de l'expression (82) que

$$\dot{q}(t)=-\left(\frac{\varepsilon^2}{\omega}+\omega\right)ae^{-\varepsilon t}\sin\omega t. \quad (8.84)$$

Ainsi l'expression (82) détermine la loi du mouvement du système dans l'intervalle de temps $0 < t < t_1$, où

$$t_1=\frac{\pi}{\omega}, \quad (8.85)$$

puisque, d'après (84), dans cet intervalle $\dot{q}(t) < 0$, ce qu'on a supposé en recourant à l'équation (81).

D'après (82), à l'instant $t = t_1$ les fonctions $q(t)$ et $\dot{q}(t)$ prennent les valeurs suivantes:

$$q(t_1)=-\beta a, \quad \dot{q}(t_1)=0, \quad (8.86)$$

avec

$$\beta=e^{-\varepsilon\frac{\pi}{\omega}}<1. \quad (8.87)$$

En adoptant les notations

$$b = \beta a, \quad (8.88)$$

on obtient d'après (86)

$$q(t_1) = -b. \tag{8.89}$$

Dans l'intervalle de temps qui suit t_1, $\dot{q}(t)$ est positif et, d'après (79), dans cet intervalle on a l'équation

$$\ddot{q} + 2\varepsilon\dot{q} + k^2 q = k^2 S, \tag{8.90}$$

qu'il faut intégrer pour les conditions initiales

$$q = -b < 0, \ \dot{q} = 0 \ \text{pour} \ t = t_1 = \frac{\pi}{\omega}. \tag{8.91}$$

Voici la solution de l'équation différentielle (90) vérifiant les conditions initiales (91)

$$q(t) = S - (b + S)\, e^{-\varepsilon(t-t_1)} \cos\omega(t - t_1) - \\ - \frac{\varepsilon}{\omega}(b + S)\, e^{-\varepsilon(t-t_1)} \sin\omega(t - t_1). \tag{8.92}$$

Il résulte de l'expression (92) que

$$\dot{q}(t) = \left(\frac{\varepsilon^2}{\omega} + \omega\right)(b + S)\, e^{-\varepsilon(t-t_1)} \sin\omega(t - t_1). \tag{8.93}$$

(93) entraîne que $\dot{q} > 0$ dans l'intervalle de temps $t_1 < t < t_2$ où

$$t_2 = \frac{2\pi}{\omega}, \tag{8.94}$$

ce que nous avons avancé en appliquant l'équation (90). Ainsi l'expression (92) définit la loi de variation de q dans l'intervalle de temps $t_1 < t < t_2$.

A l'instant $t = t_2$, $\dot{q}(t)$ s'annule

$$\dot{q}(t_2) = 0. \tag{8.95}$$

D'après (92), la valeur de $q(t)$ à l'instant $t = t_2$ s'écrit

$$q(t_2) = c, \tag{8.96}$$

où

$$c = \beta b + (1 + \beta) S. \tag{8.97}$$

Si l'on pose que q et $\dot{q}$ sont les coordonnées de phase du système, on peut construire dans le plan de phase (fig. 8.11) la trajectoire du point représentatif, déterminée par les expressions (82), (84), (92) et (93).

Nous avons examiné un cycle du mouvement du point représentatif dans le plan de phase (fig. 8.11) et établi les relations (88) et (97) entre les amplitudes consécutives a, b et c de ce cycle. On voit aisément de ce qui a été énoncé ci-dessus que ces relations ont lieu également pour les cycles qui suivent, si, certes, par a, b, c on entend

les amplitudes correspondantes du cycle successif considéré. En introduisant les notations

$$\left.\begin{aligned} f_1(a, b) &= b - \beta a, \\ f_2(c, b) &= \beta b - c + (1 + \beta) S, \end{aligned}\right\} \tag{8.98}$$

les relations (88) et (97) peuvent se mettre sous la forme

$$\left.\begin{aligned} f_1(a, b) &= 0, \\ f_2(c, b) &= 0. \end{aligned}\right\} \tag{8.99}$$

Les relations (99) peuvent être considérées comme les formules de la transformation ponctuelle qui transforment tout point a_i

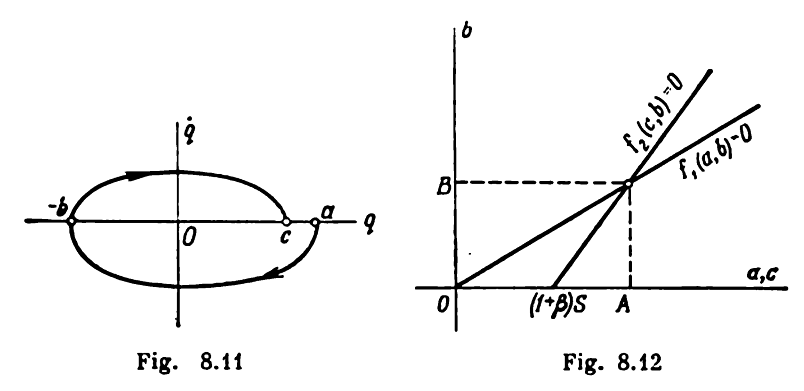


Fig. 8.11 Fig. 8.12

de l'axe des abscisses du plan de phase en point c_i de ce même axe, où i est le numéro du cycle décrit par les expressions (82) et (92). Les équations (99) peuvent également être considérées comme celles des droites dans les plans ab et cb respectifs (fig. 8.12). Si l'on fait coïncider les plans ab et cb, ces droites se coupent au point dont les coordonnées sont

$$a = c = A, \quad b = B. \tag{8.100}$$

Les valeurs de A et de B déterminées d'après les équations

$$\left.\begin{aligned} B - \beta A &= 0, \\ \beta B - A + (1 + \beta) S &= 0, \end{aligned}\right\} \tag{8.101}$$

sont

$$A = \frac{S}{1-\beta}, \quad B = \frac{\beta S}{1-\beta}. \tag{8.102}$$

Le point $q = A$ de l'axe des abscisses est le *point invariant* de la transformation ponctuelle déterminée par les relations (99). Si $a = A$, en suivant la trajectoire de phase, le point représentatif,

une fois le cycle réalisé, revient en $q = A$ (puisque $c = A$), c'est-à-dire décrit dans le plan de phase une trajectoire *fermée* (fig. 8.13). Ensuite, le point représentatif se déplace en décrivant la même trajectoire de phase, c'est-à-dire la variation de la fonction $q(t)$ est régie par une loi périodique. Notons que les grandeurs A et B ne dépendent que des paramètres du système et non des *conditions*

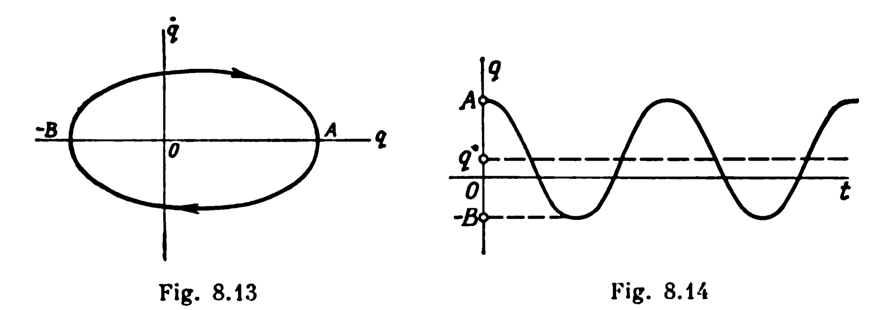


Fig. 8.13 Fig. 8.14

initiales du mouvement. C'est là une différence capitale entre le mouvement périodique que nous avons décrit et les mouvements de libration d'un système conservatif.

La courbe de la fonction $q(t)$ associée à une trajectoire de phase fermée (fig. 8.13) est représentée figure 8.14. Puisque d'après (102) $B < A$, les oscillations sont centrées non pas sur le point $q = 0$ mais sur un certain point $q^* > 0$. En amenant à la grille du tube (fig. 8.8) une tension continue supplémentaire (polarisation), on peut obtenir une trajectoire de phase fermée symétrique par rapport à l'axe des ordonnées du plan de phase.

Etudions maintenant le mouvement du point représentatif dans le voisinage d'une trajectoire de phase fermée. Soit

$$a = A + \delta a.$$

Alors, il vient

$$b = B + \delta b, \qquad c = A + \delta c.$$

Les relations (99) qui relient les amplitudes successives deviennent

$$\left.\begin{aligned} f_1(A + \delta a, B + \delta b) = 0, \\ f_2(A + \delta c, B + \delta b) = 0. \end{aligned}\right\} \qquad (8.103)$$

En développant les premiers membres des relations (103) en séries de Taylor dans le voisinage du point (A, B) on a

$$\left.\begin{aligned} f_1(A, B) + \left(\frac{\partial f_1}{\partial a}\right)_{\substack{a=A\\ b=B}} \delta a + \left(\frac{\partial f_1}{\partial b}\right)_{\substack{a=A\\ b=B}} \delta b + \ldots = 0, \\ f_2(A, B) + \left(\frac{\partial f_2}{\partial c}\right)_{\substack{c=A\\ b=B}} \delta c + \left(\frac{\partial f_2}{\partial b}\right)_{\substack{c=A\\ b=B}} \delta b + \ldots = 0. \end{aligned}\right\} \qquad (8.104)$$

En se bornant aux membres du premier ordre par rapport aux δa, δb et δc, et en tenant compte que d'après (101)

$$f_1(A, B) = 0, \quad f_2(A, B) = 0, \tag{8.105}$$

on obtient les relations suivantes:

$$\left.\begin{aligned} \left(\frac{\partial f_1}{\partial a}\right)_{\substack{a=A\\ b=B}} \delta a + \left(\frac{\partial f_1}{\partial b}\right)_{\substack{a=A\\ b=B}} \delta b = 0, \\ \left(\frac{\partial f_2}{\partial c}\right)_{\substack{c=A\\ b=B}} \delta c + \left(\frac{\partial f_2}{\partial b}\right)_{\substack{c=A\\ b=B}} \delta b = 0. \end{aligned}\right\} \tag{8.106}$$

Puisque d'après (98)

$$\frac{\partial f_1}{\partial a} = -\beta, \quad \frac{\partial f_1}{\partial b} = 1, \quad \frac{\partial f_2}{\partial c} = -1, \quad \frac{\partial f_2}{\partial b} = \beta, \tag{8.107}$$

les relations (106) deviennent

$$\left.\begin{aligned} -\beta\delta a + \delta b = 0, \\ -\delta c + \beta\delta b = 0. \end{aligned}\right\} \tag{8.108}$$

D'où

$$\delta c = \beta^2 \delta a. \tag{8.109}$$

On voit de la figure 8.11 que si par a_i on entend l'amplitude initiale du i-ième cycle, c_i sera alors l'amplitude initiale du $(i+1)$-ième cycle, c'est-à-dire

$$c_i = a_{i+1}. \tag{8.110}$$

On en déduit que

$$\delta c_i = \delta a_{i+1}. \tag{8.111}$$

Par suite, la relation (109) peut être récrite

$$\delta a_{i+1} = \beta^2 \delta a_i. \tag{8.112}$$

La relation (112) est une équation aux différences finies linéaire par rapport à la variation de l'amplitude δa. Le numéro du cycle i constitue la variable discrète de la fonction δa recherchée.

Comme d'après (87) $\beta < 1$, il vient

$$\lim_{i \to \infty} \delta a_i = 0,$$

d'où l'on tire que les trajectoires de phase s'approchent asymptotiquement de la trajectoire de phase fermée déterminée par les grandeurs (102) et représentée figure 8.13. La trajectoire de phase considérée présente ainsi un *cycle limite stable*.

Ce cycle limite détermine précisément les *auto-oscillations* générées par un oscillateur à tube.

§ 9. Systèmes non linéaires soumis à des forces périodiques extérieures

1. Oscillations forcées d'un système non linéaire. Considérons un système non linéaire à un degré de liberté dont la masse réduite est

$$m(q) = m = \text{const.} \tag{9.1}$$

Dans ce cas les équations canoniques du mouvement (8.9) s'écrivent

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{p}{m}, \\ \frac{dp}{dt} &= -V'(q) + Q(q, p, t). \end{aligned}\right\} \tag{9.2}$$

Soit

$$-V'(q) + Q(q, p, t) = m[-f(q, \dot{q}) + \psi(\nu t)], \tag{9.3}$$

où $\psi(\nu t)$ est une fonction périodique de période

$$P = \frac{2\pi}{\nu}. \tag{9.4}$$

Le système d'équations différentielles (2) peut être remplacé par une équation du second ordre

$$\ddot{q} + f(q, \dot{q}) = \psi(\nu t). \tag{9.5}$$

Introduisons maintenant les notations

$$\vartheta = \nu t. \tag{9.6}$$

Comme l'intervalle de temps t égal à une période de la fonction $\psi(\nu t)$ fait $\frac{2\pi}{\nu}$, et d'après (6), $\vartheta = \nu t$, la période de la fonction $\psi(\vartheta)$ par rapport à la nouvelle variable ϑ est égale à 2π. On sait qu'en respectant certaines conditions (par exemple, les conditions de Dirichlet), la fonction $\psi(\vartheta)$ se développe en série de Fourier

$$\psi(\vartheta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\psi(\vartheta)\,d\vartheta + \left[\frac{1}{\pi}\int_{-\pi}^{\pi}\psi(\vartheta)\cos\vartheta\,d\vartheta\right]\cos\vartheta + \\ + \left[\frac{1}{\pi}\int_{-\pi}^{\pi}\psi(\vartheta)\sin\vartheta\,d\vartheta\right]\sin\vartheta + \ldots \tag{9.7}$$

La série (7) peut être mise sous la forme:

$$\psi(\vartheta) = C + R\cos(\vartheta + \gamma) + \ldots, \tag{9.8}$$

avec

$$\left.\begin{aligned} C &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(\vartheta)\, d\vartheta, \\ R &= \sqrt{\left[\frac{1}{\pi}\int_{-\pi}^{\pi} \psi(\vartheta)\cos\vartheta\, d\vartheta\right]^2 + \left[\frac{1}{\pi}\int_{-\pi}^{\pi} \psi(\vartheta)\sin\vartheta\, d\vartheta\right]^2}, \end{aligned}\right\} \quad (9.9)$$

alors que l'angle γ est défini par les relations

$$\cos\gamma = \frac{1}{\pi R}\int_{-\pi}^{\pi} \psi(\vartheta)\cos\vartheta\, d\vartheta, \quad \sin\gamma = -\frac{1}{\pi R}\int_{-\pi}^{\pi} \psi(\vartheta)\sin\vartheta\, d\vartheta. \quad (9.10)$$

Considérons ensuite que la fonction $\psi(\vartheta)$ est symétrique, ce qui veut dire que

$$C = 0, \quad (9.11)$$

et, par suite,

$$\psi(\vartheta) = R\cos(\vartheta + \gamma) + \ldots \quad (9.12)$$

Examinons d'abord un cas *particulier*. Soit

$$\left.\begin{aligned} f(q, \dot{q}) &= (\omega^2 + \varepsilon^2)\, q + 2\varepsilon\dot{q}, \\ \psi(\nu t) &= R\cos(\vartheta + \gamma). \end{aligned}\right\} \quad (9.13)$$

L'équation (5) devient alors

$$\ddot{q} + 2\varepsilon\dot{q} + (\omega^2 + \varepsilon^2)\, q = R\cos(\nu t + \gamma). \quad (9.14)$$

La solution particulière de l'équation (14) qui correspond à son second membre est:

$$q = a\cos(\nu t + \gamma - \theta). \quad (9.15)$$

Comme d'après (2) $p = m\dot{q}$, dans le cas envisagé

$$p = -ma\nu\sin(\nu t + \gamma - \theta). \quad (9.16)$$

Ici a est l'amplitude des oscillations forcées et θ, leur déphasage par rapport à la phase de la force extérieure $\psi(\nu t)$. L'équation (14) étant une équation linéaire, dans les expressions (15) et (16) $a = \text{const}$ et $\theta = \text{const}$.

D'après (2) et (3), dans le cas *général* les équations du mouvement s'écrivent

$$\left.\begin{aligned} \frac{dq}{dt} &= \frac{p}{m}, \\ \frac{1}{m}\frac{dp}{dt} + f\left(q, \frac{p}{m}\right) &= \psi(\vartheta), \end{aligned}\right\} \quad (9.17)$$

et les expressions

$$q = a\cos(\vartheta + \gamma - \theta), \quad p = -ma\nu\sin(\vartheta + \gamma - \theta), \tag{9.18}$$

où $a = \text{const}$, $\theta = \text{const}$, ne vérifient pas les équations (17).

Pour cette raison dans le cas général, nous considérons les expressions (18) comme des formules de changement des variables en posant que a et θ sont de nouvelles variables, c'est-à-dire en posant dans les formules (18)

$$a = a(t), \qquad \theta = \theta(t). \tag{9.19}$$

Aussi l'écriture de la solution particulière (15), (16) de l'équation linéaire (14) n'est-elle utilisée ici que pour guider le choix des formules de changement des variables.

Au lieu de q et p, portons maintenant dans les équations (17) les expressions (18) que nous écrirons

$$\left.\begin{aligned} q &= a\cos u, \\ p &= -ma\nu\sin u, \end{aligned}\right\} \tag{9.20}$$

où

$$u = \nu t + \gamma - \theta = \vartheta + \gamma - \theta. \tag{9.21}$$

Remarquons que d'après (10), $\gamma = \text{const}$ et il vient

$$\dot{u} = \nu - \dot{\theta}. \tag{9.22}$$

Ainsi, les équations (17) se ramènent à la forme

$$\left.\begin{aligned} \dot{a}\cos u - a(\nu - \dot{\theta})\sin u &= -a\nu\sin u, \\ -\dot{a}\nu\sin u - a\nu(\nu - \dot{\theta})\cos u + f(a\cos u, \ -a\nu\sin u) &= \psi(\vartheta). \end{aligned}\right\} \tag{9.23}$$

Pour résoudre les équations par rapport à $\dot{a}$ et $\dot{\theta}$, multiplions la première équation (23) par $\cos u$, et la deuxième par $-\frac{1}{\nu}\sin u$ et additionnons respectivement les premiers et les seconds membres des équations ainsi obtenues. Il en résulte une équation dans laquelle $\dot{\theta}$ ne figure plus. En multipliant la première équation (23) par $\sin u$ et la deuxième équation (23) par $\frac{1}{\nu}\cos u$, puis en additionnant les équations obtenues, on trouve une équation qui ne possède pas de $\dot{a}$.

Il en résulte donc les équations suivantes

$$\left.\begin{aligned} \dot{a} &= -a\nu\sin u\cos u + \frac{1}{\nu} f(a\cos u, \ -a\nu\sin u)\sin u - \\ &\qquad\qquad - \frac{1}{\nu}\psi(\vartheta)\sin u, \\ \dot{\theta} &= \nu\cos^2 u - \frac{1}{\nu a} f(a\cos u, \ -a\nu\sin u)\cos u + \\ &\qquad\qquad + \frac{1}{\nu a}\psi(\vartheta)\cos u. \end{aligned}\right\} \tag{9.24}$$

Les équations (24) sont des équations *exactes*. Pour ramener les équations initiales aux nouvelles variables nous n'avons rien simplifié. L'intégration des équations (24) n'est donc point plus facile que celle des équations initiales. Dans le cas général on ne parvient pas à intégrer les équations (24) et à cet effet on est obligé de faire appel à des méthodes approchées.

Si l'on se borne aux fonctions $f\left(q, \frac{p}{m}\right)$, *peu différentes des fonctions linéaires*, le système d'équations (24) peut être approximé par des équations plus simples, en s'appuyant sur le fait que si les fonctions $f\left(q, \frac{p}{m}\right)$ qui interviennent dans (24) sont voisines des fonctions linéaires, les fonctions cherchées a et θ varient lentement dans le temps. Cette propriété des systèmes dits systèmes pseudo-linéaires est due au fait que la solution dépend de façon continue du paramètre, car si les fonctions $f\left(q, \frac{p}{m}\right)$ sont linéaires, alors $a = \text{const}$ et $\theta = \text{const}$.

Les seconds membres des équations (24) sont des fonctions *périodiques par rapport à la variable u* et se prêtent au développement en séries de Fourier. Etant donné que a et θ des systèmes pseudo-linéaires sont des fonctions du temps à *variation lente*, les solutions de (24) sont influencées surtout par les termes *constants* des séries de Fourier; l'action exercée par les termes *oscillatoires*, c'est-à-dire contenant le premier harmonique et des harmoniques supérieurs, est assez faible.

Voici un exemple à titre d'illustration. La solution de l'équation différentielle

$$\dot{a} + ka = S + A \sin \nu t$$

s'écrit

$$a(t) = \left[a(0) - \frac{S}{k} + \frac{A\nu}{k^2+\nu^2}\right] e^{-kt} + \frac{S}{k} + \frac{Ak}{k^2+\nu^2} \sin \nu t - \frac{A\nu}{k^2+\nu^2} \cos \nu t.$$

Si $\nu \gg k$, il vient

$$a(t) \approx \left[a(0) - \frac{S}{k} + \frac{A}{\nu}\right] e^{-kt} + \frac{S}{k} - \frac{A}{\nu}\left(\cos \nu t - \frac{k}{\nu} \sin \nu t\right).$$

Puisque pour $\nu \gg k$ on a l'inégalité $\frac{1}{k} \gg \frac{1}{\nu}$, la partie principale de la solution est

$$a(t) \approx \left[a(0) - \frac{S}{k}\right] e^{-kt} + \frac{S}{k}.$$

La fonction $A \sin \nu t$ rapidement oscillante influe relativement peu sur la solution de l'équation étudiée.

Compte tenu de cette circonstance, adoptons en première approximation les équations qui s'obtiennent en remplaçant dans (24) les seconds membres par leurs valeurs moyennes par rapport à u,

c'est-à-dire par les termes constants de leurs développements en séries de Fourier. On a ainsi les équations de la première approximation

$$\left.\begin{aligned}\dot{a}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\Big[-a\nu\sin u\cos u+&\\ +\frac{1}{\nu}f(a\cos u,\ -a\nu\sin u)\sin u-\frac{1}{\nu}\psi(\vartheta)\sin u\Big]\,du,&\\ \dot{\theta}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\Big[\nu\cos^2 u-&\\ -\frac{1}{\nu a}f(a\cos u,\ -a\nu\sin u)\cos u+\frac{1}{\nu a}\psi(\vartheta)\cos u\Big]\,du.&\end{aligned}\right\}\qquad(9.25)$$

Pour calculer les intégrales des seconds membres de (25) considérons que a et θ sont constants, c'est-à-dire nous négligeons la variation de ces fonctions en une période. L'ordre de petitesse de l'erreur ainsi introduite sera déjà plus élevé.

Comme

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\sin u\cos u\,du=0,\quad \frac{1}{2\pi}\int_{-\pi}^{\pi}\nu\cos^2 u\,du=\frac{\nu}{2},\qquad(9.26)$$

l'équation (25) peut se mettre sous la forme

$$\left.\begin{aligned}\dot{a}=\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}f(a\cos u,\ -a\nu\sin u)\sin u\,du-&\\ -\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}\psi(\vartheta)\sin u\,du,&\\ \dot{\theta}=\frac{\nu}{2}-\frac{1}{2\pi\nu a}\int_{-\pi}^{\pi}f(a\cos u,\ -a\nu\sin u)\cos u\,du+&\\ +\frac{1}{2\pi\nu a}\int_{-\pi}^{\pi}\psi(\vartheta)\cos u\,du.&\end{aligned}\right\}\qquad(9.27)$$

Notons maintenant que

$$\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}\psi(\vartheta)\sin u\,du=\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}\psi(\vartheta)\sin(\vartheta+\gamma-\theta)\,d\vartheta=$$

$$=\left[\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}\psi(\vartheta)\sin\vartheta\,d\vartheta\right]\cos(\gamma-\theta)+$$

$$+\left[\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}\psi(\vartheta)\cos\vartheta\, d\vartheta\right]\sin(\gamma-\theta)=$$

$$=\frac{1}{2\nu}\left[-R\sin\gamma\cos(\gamma-\theta)+R\cos\gamma\sin(\gamma-\theta)\right]=-\frac{R}{2\nu}\sin\theta. \quad (9.28)$$

D'une façon analogue

$$\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}\psi(\vartheta)\cos u\, du=\frac{R}{2\nu}\cos\theta. \quad (9.29)$$

En portant les expressions (28) et (29) dans l'équation (27) on a

$$\left.\begin{aligned}\dot{a}&=\frac{1}{2\pi\nu}\int_{-\pi}^{\pi}f(a\cos u,\ -a\nu\sin u)\sin u\, du+\frac{R}{2\nu}\sin\theta,\\ \dot{\theta}&=\frac{\nu}{2}-\frac{1}{2\pi\nu a}\int_{-\pi}^{\pi}f(a\cos u,\ -a\nu\sin u)\cos u\, du+\frac{R}{2\nu a}\cos\theta.\end{aligned}\right\} \quad (9.30)$$

Introduisons maintenant les notations

$$\left.\begin{aligned}\Phi(\nu, a)&=-\frac{1}{2\pi\nu a}\int_{-\pi}^{\pi}f(a\cos u,\ -a\nu\sin u)\sin u\, du,\\ \Psi(\nu, a)&=\frac{\nu}{2}+\frac{1}{2\pi\nu a}\int_{-\pi}^{\pi}f(a\cos u,\ -a\nu\sin u)\cos u\, du.\end{aligned}\right\} \quad (9.31)$$

Les équations (30) deviennent alors

$$\left.\begin{aligned}\dot{a}&=-a\Phi(\nu, a)+\frac{R}{2\nu}\sin\theta,\\ \dot{\theta}&=\nu-\Psi(\nu, a)+\frac{R}{2\nu a}\cos\theta.\end{aligned}\right\} \quad (9.32)$$

Comme nous l'avons déjà dit, les équations (32) sont des équations de première approximation. On les appelle équations *raccourcies*. La méthode par laquelle on les a obtenues s'appelle *méthode de la moyenne*.

La construction des approximations plus élevées et la définition des conditions de convergence des approximations pour les systèmes à n degrés de liberté sont établies par B. Boulgakov dans sa monographie [17] qui a guidé l'auteur dans l'exposé des problèmes des §§ 8 et 9.

Le développement et la justification des méthodes asymptotiques de la théorie des oscillations non linéaires, auxquelles se rapporte également la méthode de la moyenne, font entre autres, l'objet des travaux de B. Van der Pol, L. Mandelstam, A. Andronov, N. Krylov, N. Bogolioubov, Y. Mitropolski [4, 12, 47, 64].

2. Oscillations stationnaires à fréquence de la force extérieure et leur stabilité. La solution particulière des équations (32)

$$\left.\begin{aligned} a &= a^* = \text{const},\\ \theta &= \theta^* = \text{const}\end{aligned}\right\} \tag{9.33}$$

par rapport aux variables initiales correspond d'après (20) à la solution suivante

$$q = a^* \cos(\nu t + \gamma - \theta^*). \tag{9.34}$$

La solution (34) présente les *oscillations forcées* du système, c'est-à-dire les oscillations stationnaires à fréquence ν de la force extérieure.

Pour trouver les valeurs a^* et θ^*, déterminons d'après (32) le système suivant des équations finies (en général, transcendantes):

$$\left.\begin{aligned} -a\Phi(\nu, a) + \frac{R}{2\nu}\sin\theta = 0,\\ \nu - \Psi(\nu, a) + \frac{R}{2\nu a}\cos\theta = 0.\end{aligned}\right\} \tag{9.35}$$

Les racines des équations (35) sont précisément les valeurs a^* et θ^* qui nous intéressent.

En éliminant θ du système d'équations (35) on obtient l'équation suivante par rapport à a:

$$[a\Phi(\nu, a)]^2 + [\nu a - a\Psi(\nu, a)]^2 = \left(\frac{R}{2\nu}\right)^2. \tag{9.36}$$

Si l'on désigne

$$W(\nu, a) = [a\Phi(\nu, a)]^2 + [\nu a - a\Psi(\nu, a)]^2 - \left(\frac{R}{2\nu}\right)^2, \tag{9.37}$$

l'équation (36) devient

$$W(\nu, a) = 0. \tag{9.38}$$

Pour chaque valeur donnée de ν, l'équation (38) permet de trouver r valeurs (où r est le nombre de racines de l'équation) $a_i = a_i^*$ $(i = 1, \ldots, r)$ des amplitudes des oscillations forcées possibles lorsque le système subit l'action d'une force extérieure de période $2\pi/\nu$.

Pour toute valeur a_i^* l'équation (35) permet d'obtenir les valeurs $\sin\theta^*$ et $\cos\theta^*$ qui lui correspondent et suivant lesquelles on détermine la valeur θ_i^*.

L'équation $W(\nu, a) = 0$ définit dans le plan νa la courbe $a = a^*(\nu)$ caractéristique de la relation entre l'amplitude des oscillations forcées a^* et la fréquence ν de la force extérieure. Cette courbe s'appelle *courbe de résonance*.

Certains types possibles de courbes de résonance sont représentés figure 9.1. L'expression (37) montre qu'aux points de l'axe

des abscisses ($a = 0$) la fonction $W(\nu, a)$ prend des valeurs négatives

$$W(\nu, 0) = -\left(\frac{R}{2\nu}\right)^2 < 0.$$

Sur la courbe de résonance la fonction $W(\nu, a)$ s'annule. L'allure continue conduit à la conclusion que la fonction $W(\nu, a)$ est négative dans la zone comprise entre l'axe des abscisses et la courbe

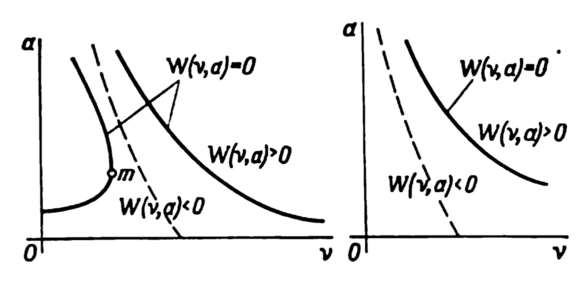


Fig. 9.1

de résonance et positive en dehors de cette zone, comme on le voit sur la figure 9.1.

La relation (38) implique qu'aux points de la courbe de résonance

$$\frac{dW}{d\nu} = \frac{\partial W}{\partial \nu} + \frac{\partial W}{\partial a}\frac{da}{d\nu} = 0. \tag{9.39}$$

Indiquons maintenant les points où $\frac{\partial W}{\partial a}$ s'annule.

1) Soit en un point $\frac{\partial W}{\partial \nu} = 0$, $\frac{da}{d\nu} \neq 0$. Alors, en ce point, en vertu de (39), $\frac{\partial W}{\partial a} = 0$. Ainsi, c'est le point où $\frac{\partial W}{\partial \nu} = 0$, $\frac{\partial W}{\partial a} = 0$, c'est-à-dire c'est un point singulier.

2) Soit en un point $\frac{\partial W}{\partial \nu} \neq 0$. Si en ce point $\frac{da}{d\nu} = \infty$, on a en vertu de (39) $\frac{\partial W}{\partial a} = 0$. Par conséquent, nous avons là un point de la courbe de résonance où la tangente est *verticale*. (Sur la figure 9.1 un tel point est m.)

En tout autre point de la courbe de résonance, $\frac{\partial W}{\partial a} \neq 0$. De plus, comme il s'ensuit de la disposition relative des zones positive et négative de la fonction $W(\nu, a)$, $\frac{\partial W}{\partial a} < 0$ seulement aux points de l'arc situé au-dessus du point m (fig. 9.1) sur la branche gauche de la courbe de résonance. En d'autres points de la courbe de résonance, $\frac{\partial W}{\partial a} > 0$.

Nous avons déjà dit que les mouvements stationnaires à fréquence de la force extérieure (c'est-à-dire les oscillations forcées) sont déterminés par la solution particulière (33) des équations (32). Examinons maintenant la *stabilité* de ces mouvements.

Soient

$$a = a^* + \delta a, \qquad \theta = \theta^* + \delta\theta, \tag{9.40}$$

où a^* et θ^* vérifient les équations (35).

En portant les grandeurs (40) dans les équations (32) et en développant les seconds membres de ces équations en séries de Taylor dans le voisinage du point (a^*, θ^*), on a

$$\left.\begin{aligned} \frac{d}{dt}\delta a &= -a^*\Phi(\nu, a^*) + \frac{R}{2\nu}\sin\theta^* - \\ &\qquad - \left[\frac{\partial}{\partial a} a\Phi(\nu, a)\right]_{a=a^*}\delta a + \frac{R}{2\nu}\cos\theta^*\delta\theta + \ldots, \\ \frac{d}{dt}\delta\theta &= \nu - \Psi(\nu, a^*) + \frac{R}{2\nu a^*}\cos\theta^* - \left[\frac{\partial}{\partial a}\Psi(\nu, a)\right]_{a=a^*}\delta a - \\ &\qquad - \frac{R}{2\nu a^{*2}}\cos\theta^*\delta a - \frac{R}{2\nu a^*}\sin\theta^*\delta\theta + \ldots \end{aligned}\right\} \tag{9.41}$$

D'après (35)

$$\left.\begin{aligned} a^*\Phi(\nu, a^*) &= \frac{R}{2\nu}\sin\theta^*, \\ -[\nu a^* - a^*\Psi(\nu, a^*)] &= \frac{R}{2\nu}\cos\theta^*. \end{aligned}\right\} \tag{9.42}$$

En tenant compte des relations (42) et en ne gardant dans les équations (41) que les termes du premier ordre en δa et $\delta\theta$, on obtient les *équations aux variations* suivantes:

$$\left.\begin{aligned} &\frac{d}{dt}\delta a + \left[\frac{\partial}{\partial a} a\Phi(\nu, a)\right]_{a=a^*}\delta a - \frac{R}{2\nu}\cos\theta^*\delta\theta = 0, \\ &\frac{d}{dt}\delta\theta + \left[\frac{\partial}{\partial a}\Psi(\nu, a)\right]_{a=a^*}\delta a + \frac{R}{2\nu a^{*2}}\cos\theta^*\delta a + \\ &\qquad\qquad + \frac{R}{2\nu a^*}\sin\theta^*\delta\theta = 0. \end{aligned}\right\} \tag{9.43}$$

Les équations (43) sont des équations différentielles linéaires aux coefficients constants. En y remplaçant $\frac{R}{2\nu}\sin\theta^*$ et $\frac{R}{2\nu}\cos\theta^*$ par leurs valeurs tirées de (42), on peut ramener les équations (43) à la forme

$$\left.\begin{aligned} &\frac{d}{dt}\delta a + \left[\frac{\partial}{\partial a} a\Phi(\nu, a)\right]_{a=a^*}\delta a + [\nu a^* - a^*\Psi(\nu, a^*)]\,\delta\theta = 0, \\ &\frac{d}{dt}\delta\theta + \left[\frac{\partial}{\partial a}\Psi(\nu, a)\right]_{a=a^*}\delta a - \frac{1}{a^*}[\nu - \Psi(\nu, a^*)]\,\delta a + \\ &\qquad\qquad + \Phi(\nu, a^*)\,\delta\theta = 0. \end{aligned}\right\} \tag{9.44}$$

L'équation caractéristique du système d'équations différentielles (44) est

$$\lambda^2 + M_1\lambda + M_2 = 0, \tag{9.45}$$

où

$$M_1 = \Phi(\nu, a^*) + \left[\frac{\partial}{\partial a} a\Phi(\nu, a)\right]_{a=a^*}, \tag{9.46}$$

$$M_2 = \Phi(\nu, a^*) \left[\frac{\partial}{\partial a} a\Phi(\nu, a)\right]_{a=a^*} - \\ - [\nu a^* - a^*\Psi(\nu, a^*)] \left[\frac{\partial}{\partial a} \Psi(\nu, a)\right]_{a=a^*} + [\nu - \Psi(\nu, a^*)]^2. \tag{9.47}$$

Les mouvements stationnaires considérés (c'est-à-dire les oscillations forcées) sont asymptotiquement stables si δa et $\delta\theta$, déterminées par les équations (44), tendent asymptotiquement vers zéro quand $t \to \infty$. Ceci a lieu si les racines de (45) satisfont à la condition

$$\mathrm{Re}\, \lambda_1 < 0, \qquad \mathrm{Re}\, \lambda_2 < 0. \tag{9.48}$$

Quant à la condition (48), pour qu'elle ait lieu, il faut et il suffit de respecter les conditions

$$M_1 > 0, \qquad M_2 > 0. \tag{9.49}$$

Si au moins un des coefficients M_1 ou M_2 est négatif, le mouvement stationnaire considéré sera instable.

L'observation de la condition $M_1 > 0$ peut être vérifiée en calculant M_1 d'après la formule (46).

Pour ce qui est de la condition $M_2 > 0$, on peut en juger directement d'après la courbe de résonance. Il en est ainsi du fait que d'après (37) et (47)

$$M_2 = \left(\frac{1}{2a} \frac{\partial W}{\partial a}\right)_{a=a^*}. \tag{9.50}$$

Ainsi le signe de M_2 coïncide avec celui de $\frac{\partial W}{\partial a}$ en $a = a^*$. Pour ce qui est du signe de $\frac{\partial W}{\partial a}$, il est déterminé suivant l'allure de la courbe de résonance, comme nous l'avons décrit précédemment. En particulier, il résulte de la relation (50) que tout point de l'arc situé au-dessus du point m de la branche gauche de la courbe de résonance (fig. 9.1) correspond aux oscillations forcées *instables*, parce qu'en n'importe lequel de ces points $\frac{\partial W}{\partial a} < 0$.

CHAPITRE 3

SYSTÈMES À COMMANDE EN TEMPS FINI

§ 10. Fonctions des matrices et leur application à l'intégration des systèmes d'équations différentielles linéaires

L'application des fonctions des matrices à la théorie des systèmes d'équations différentielles linéaires a fait l'objet des ouvrages fondamentaux de I. Lappo-Danilevski [50]. Ces problèmes ont également été étudiés avec force détails dans les monographies de F. Gantmacher [21] et B. Boulgakov [17]. Ce sont des ouvrages qui ont guidé l'auteur dans l'exposé qui suit. Les méthodes de la théorie des fonctions des matrices trouvent leur application directe dans les problèmes de commandabilité et d'observabilité étudiés ci-dessous.

1. Théorème de Bezout généralisé. Soit $A = [A_{jk}]$ une matrice carrée $n \times n$. Désignons par $F(\lambda) = [F_{jk}(\lambda)]$ $(n \times n)$ le polynôme matriciel de degré m

$$F(\lambda) = F_0\lambda^m + F_1\lambda^{m-1} + \ldots + F_m \quad (F_0 \neq 0). \qquad (10.1)$$

Constatons que si dans l'expression (1) le scalaire λ est remplacé par la matrice A, on a

$$F(A) = F_0A^m + F_1A^{m-1} + \ldots + F_mE, \qquad (10.2)$$

$$\hat{F}(A) = A^mF_0 + A^{m-1}F_1 + \ldots + EF_m, \qquad (10.3)$$

et de plus, dans le cas général, $F(A) \neq \hat{F}(A)$.

Divisons le polynôme matriciel $F(\lambda)$ par le binôme $\lambda E - A$ à droite et à gauche

$$F(\lambda) = Q(\lambda)(\lambda E - A) + R, \quad F(\lambda) = (\lambda E - A)\hat{Q}(\lambda) + \hat{R}. \quad (10.4)$$

Comme après la division des polynômes le degré du reste est inférieur à celui du diviseur, le reste droit R et le reste gauche $\hat{R}$ ne dépendent pas de λ.

En substituant dans les identités (4) la matrice A à λ, opération possible du fait que la multiplication de la matrice A par les coefficients matriciels du binôme $\lambda E - A$ étant commutative, on obtient

$$F(A) = Q(A)(A - A) + R, \quad \hat{F}(A) = (A - A)\hat{Q}(A) + \hat{R}. \quad (10.5)$$

On en déduit le théorème suivant.

Théorème de Bezout généralisé. *Si l'on divise à droite (à gauche) le polynôme matriciel $F(\lambda)$ par le binôme $\lambda E - A$, le reste de la division est égal à $F(A)$ (respectivement à $\hat{F}(A)$.).*

Corollaire. *Le polynôme matriciel $F(\lambda)$ est divisible à droite (à gauche) sans reste par le binôme $\lambda E - A$ si et seulement si $F(A) = 0$ (respectivement $\hat{F}(A) = 0$).*

2. Théorème de Cayley-Hamilton. Considérons la matrice $A = [A_{jk}]$ $(n \times n)$. On appelle matrice caractéristique de A la matrice $\lambda E - A$. Le déterminant de la matrice caractéristique

$$\Delta(\lambda) = \det(\lambda E - A) \tag{10.6}$$

est un polynôme scalaire en λ dit *polynôme caractéristique* de la matrice A. Les racines de l'équation $\Delta(\lambda) = 0$ s'appellent *valeurs propres* ou *nombres caractéristiques* de A.

Désignons par $B_{jk}(\lambda)$ le cofacteur de l'élément $\lambda E_{kj} - A_{kj}$ du déterminant $\Delta(\lambda)$. La matrice $B(\lambda) = [B_{jk}]$ $(n \times n)$ est une matrice *adjointe* de la matrice $\lambda E - A$.

Les définitions précédentes conduisent aux identités suivantes par rapport à λ:

$$(\lambda E - A) B(\lambda) = \Delta(\lambda) E, \tag{10.7}$$

$$B(\lambda)(\lambda E - A) = \Delta(\lambda) E. \tag{10.8}$$

Comme

$$\Delta(\lambda) = \lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \ldots + c_{n-1}\lambda + c_n, \tag{10.9}$$

il s'ensuit que

$$\Delta(\lambda) E = E\lambda^n + (c_1E)\lambda^{n-1} + \ldots + (c_{n-1}E)\lambda + c_nE \tag{10.10}$$

est un polynôme en λ à coefficients matriciels. La matrice $B(\lambda)$ peut s'écrire également sous la forme d'un polynôme en λ à coefficients matriciels.

Les égalités (7) et (8) montrent que $\Delta(\lambda) E$ est divisible à droite et à gauche *sans reste* par $\lambda E - A$. D'après le théorème de Bezout généralisé ceci n'est possible que si le reste $\Delta(A)E = \Delta(A)$ est nul. On en tire le théorème suivant.

Théorème de Cayley-Hamilton. *Toute matrice carrée A satisfait à son équation caractéristique, c'est-à-dire*

$$\Delta(A) = 0. \tag{10.11}$$

3. Polynôme minimal associé à une matrice. Le polynôme scalaire $f(\lambda)$ s'appelle *polynôme annulateur* d'une matrice carrée A si

$$f(A) = 0.$$

Le polynôme annulateur $\psi(A)$ de *plus petit* degré et dont le coefficient supérieur est l'unité s'appelle *polynôme minimal* associé à la matrice A.

D'après le théorème de Cayley-Hamilton, le polynôme caractéristique $\Delta(\lambda)$ de la matrice A est son polynôme annulateur. Toutefois, dans le cas général, ce polynôme n'est pas minimal.

On peut montrer qu'un polynôme annulateur arbitraire $f(\lambda)$ d'une matrice se divise sans reste par son polynôme minimal $\psi(\lambda)$. En effet, en divisant $f(\lambda)$ par $\psi(\lambda)$, on a

$$f(\lambda) = \psi(\lambda)\, q(\lambda) + r(\lambda), \tag{10.12}$$

où le degré de $r(\lambda)$ est inférieur à celui de $\psi(\lambda)$. La relation (12) entraîne

$$f(A) = \psi(A)\, q(A) + r(A). \tag{10.13}$$

Puisque $f(A) = 0$, $\psi(A) = 0$, on a également $r(A) = 0$. Mais le degré de $r(\lambda)$ est inférieur à celui du polynôme minimal $\psi(\lambda)$. Donc $r(\lambda) \equiv 0$, c'est-à-dire $f(\lambda)$ se divise sans reste par $\psi(\lambda)$.

Soient $\psi(\lambda)$ et $\psi_1(\lambda)$ des polynômes minimaux associés à la matrice A. Dans ce cas, ils se divisent l'un par l'autre sans reste, c'est-à-dire ces polynômes diffèrent par un coefficient constant. Les coefficients supérieurs de $\psi(\lambda)$ et de $\psi_1(\lambda)$ étant égaux à l'unité, $\psi_1(\lambda) \equiv \psi(\lambda)$. Nous avons démontré ainsi *l'unicité* du polynôme minimal associé à la matrice A donnée.

Désignons par $D_{n-1}(\lambda)$ le plus grand commun diviseur de tous les mineurs d'ordre $(n - 1)$ d'une matrice caractéristique $\lambda E - A$, c'est-à-dire le plus grand commun diviseur des éléments de la matrice adjointe $B(\lambda)$. De plus, réduisons à l'unité le coefficient affecté au degré supérieur en λ du polynôme $D_{n-1}(\lambda)$. Appelons ce dernier *diviseur déterminant* d'ordre $(n - 1)$ de la matrice $\lambda E - A$. Ce qui vient d'être dit amène

$$B(\lambda) = D_{n-1}(\lambda) C(\lambda), \tag{10.14}$$

où $C(\lambda)$ est une matrice polynomiale dite *matrice adjointe réduite* de la matrice $\lambda E - A$.

En multipliant les premier et deuxième membres de la relation (14) à gauche par la matrice $\lambda E - A$ on a, d'après (7),

$$\Delta(\lambda)\, E = (\lambda E - A)\, C(\lambda)\, D_{n-1}(\lambda). \tag{10.15}$$

En développant le déterminant caractéristique $\Delta(\lambda)$ par rapport aux éléments d'une ligne quelconque on obtient une somme des termes dont chacun se divise sans reste par $D_{n-1}(\lambda)$, et, donc, $\Delta(\lambda)$ se divise sans reste par $D_{n-1}(\lambda)$, c'est-à-dire

$$\frac{\Delta(\lambda)}{D_{n-1}(\lambda)} = \psi(\lambda), \tag{10.16}$$

où $\psi(\lambda)$ est un polynôme. (Ce polynôme est le plus grand facteur invariant de la matrice caractéristique $\lambda E - A$.)

Les expressions (15) et (16) amènent

$$\psi(\lambda) E = (\lambda E - A) C(\lambda). \qquad (10.17)$$

Remarquons que la multiplication à droite des deux membres de la relation (14) par la matrice $\lambda E - A$ donne d'après (8)

$$\Delta(\lambda) E = D_{n-1}(\lambda) C(\lambda)(\lambda E - A), \qquad (10.18)$$

d'où, conformément à (16), on a

$$\psi(\lambda) E = C(\lambda)(\lambda E - A), \qquad (10.19)$$

c'est-à-dire $C(\lambda)$ est à la fois le quotient à gauche et le quotient à droite de la division de $\psi(\lambda) E$ par $\lambda E - A$.

Ainsi le polynôme matriciel $\psi(\lambda) E$ se divise sans reste à gauche (et à droite) par la matrice $\lambda E - A$. Par suite d'après le théorème de Bezout généralisé,

$$\psi(A) E = \psi(A) = 0, \qquad (10.20)$$

et, par conséquent, $\psi(\lambda)$ est un polynôme annulateur de la matrice A.

Montrons maintenant que $\psi(\lambda)$ est le polynôme minimal. Soit $\psi^*(\lambda)$ le polynôme minimal de la matrice A. $\psi(\lambda)$ se divise alors sans reste par $\psi^*(\lambda)$, c'est-à-dire

$$\psi(\lambda) = \psi^*(\lambda) \chi(\lambda). \qquad (10.21)$$

Puisque $\psi^*(A) = 0$, en vertu du théorème de Bezout généralisé le polynôme matriciel $\psi^*(\lambda) E$ se divise à gauche sans reste par $\lambda E - A$:

$$\psi^*(\lambda) E = (\lambda E - A) C^*(\lambda). \qquad (10.22)$$

En multipliant les deux membres de la relation (22) par le polynôme $\chi(\lambda)$, on obtient

$$\psi(\lambda) E = (\lambda E - A) C^*(\lambda) \chi(\lambda). \qquad (10.23)$$

La comparaison des relations (23) et (17) conduit à

$$C(\lambda) = C^*(\lambda) \chi(\lambda). \qquad (10.24)$$

La relation (24) entraîne que $\chi(\lambda)$ est un commun diviseur de tous les éléments de la matrice polynomiale $C(\lambda)$. Mais, comme on le voit de (14), le plus grand commun diviseur de tous les éléments de la matrice $C(\lambda)$ est égal à un certain nombre constant du fait que la matrice $C(\lambda)$ a été obtenue en divisant la matrice $B(\lambda)$ par le plus grand commun diviseur $D_{n-1}(\lambda)$ de ses éléments. Ainsi, $\chi(\lambda) = \mu = \text{const}$. Comme les coefficients dominants de $\psi(\lambda)$ et $\psi^*(\lambda)$ sont égaux à l'unité, (21) conduit à $\mu = 1$, et donc $\chi(\lambda) = 1$. Ainsi, $\psi(\lambda) = \psi^*(\lambda)$, c'est-à-dire $\psi(\lambda)$ est un polynôme minimal de la matrice A.

De cette façon, d'après (16) le polynôme minimal de la matrice A est déterminé par la formule

$$\psi(\lambda)=\frac{\Delta(\lambda)}{D_{n-1}(\lambda)}. \qquad (10.25)$$

Considérons maintenant la formule (17):

$$\psi(\lambda) E = (\lambda E - A) C(\lambda).$$

Le déterminant du produit de deux matrices étant égal au produit de leurs déterminants, et $\det [\psi(\lambda) E] = [\psi(\lambda)]^n$, on a

$$[\psi(\lambda)]^n = \Delta(\lambda) \det C(\lambda). \qquad (10.26)$$

Il s'ensuit que $[\psi(\lambda)]^n$ se divise sans reste par $\Delta(\lambda)$, et en vertu de (25), $\Delta(\lambda)$ se divise sans reste par $\psi(\lambda)$; par suite, les zéros du polynôme $\psi(\lambda)$ coïncident avec les zéros du polynôme $\Delta(\lambda)$, bien que dans le cas général les ordres de multiplicité de ces zéros diffèrent pour $\psi(\lambda)$ et $\Delta(\lambda)$. Par conséquent, *toutes les valeurs propres différentes entre elles de la matrice A sont les zéros du polynôme minimal* $\psi(\lambda)$.

On en tire que si

$$\Delta(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \dots (\lambda - \lambda_s)^{n_s}, \qquad (10.27)$$

où

$$n_1 + n_2 + \dots + n_s = n, \qquad (10.28)$$

le polynôme minimal $\psi(\lambda)$ est de la forme

$$\psi(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_s)^{m_s}, \qquad (10.29)$$

avec

$$1 \leqslant m_k \leqslant n_k \ (k = 1, 2, \dots, s). \qquad (10.30)$$

Le degré m du polynôme minimal $\psi(\lambda)$ est

$$m = m_1 + m_2 + \dots + m_s. \qquad (10.31)$$

Les binômes $(\lambda - \lambda_k)^{m_k}$ $(k = 1, \dots, s)$ sont *les plus grands diviseurs élémentaires de la matrice $\lambda E - A$*.

4. Fonctions des matrices. Soit la matrice carrée $A = [A_{jk}] (n \times n)$ et la fonction $f(\lambda)$ de la variable scalaire λ. Déterminer ce qu'il faut entendre par $f(A)$, c'est-à-dire étendre également la fonction $f(\lambda)$ aux valeurs matricielles de la variable.

Dans ce qui précède (29) nous avons désigné par $\psi(\lambda)$ le polynôme minimal de la matrice A:

$$\psi(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_s)^{m_s}$$
$$(m_1 + m_2 + \dots + m_s = m).$$

Soient deux polynômes $g(\lambda)$ et $h(\lambda)$ tels que

$$g(A) = h(A). \qquad (10.32)$$

La relation (32) entraîne que la différence de ces deux polynômes

$$d(\lambda) = g(\lambda) - h(\lambda) \tag{10.33}$$

est un polynôme annulateur de la matrice A et, par conséquent, cette différence se divise sans reste par le polynôme minimal $\psi(\lambda)$, c'est-à-dire

$$\frac{d(\lambda)}{\psi(\lambda)} = \frac{g(\lambda) - h(\lambda)}{\psi(\lambda)} = S(\lambda), \tag{10.34}$$

où $S(\lambda)$ est un polynôme en λ. Ainsi

$$g(\lambda) = h(\lambda) + S(\lambda)\,\psi(\lambda). \tag{10.35}$$

La relation (35) s'écrit

$$g(\lambda) \equiv h(\lambda) \quad (\mathrm{mod}\ \psi(\lambda)) \tag{10.36}$$

ce qui se lit: la relation $g(\lambda)$ est congrue à $h(\lambda)$ modulo $\psi(\lambda)$.

La relation (34) entraîne que

$$d(\lambda) = \psi(\lambda)\,S(\lambda), \tag{10.37}$$

c'est pourquoi, en vertu de (29), on a

$$d(\lambda_k) = 0,\ d'(\lambda_k) = 0,\ \ldots,\ d^{(m_k-1)}(\lambda_k) = 0 \quad (k = 1, \ldots, s) \tag{10.38}$$

ou d'après (33)

$$g(\lambda_k) = h(\lambda_k),\ g'(\lambda_k) = h'(\lambda_k),\ \ldots$$
$$\ldots,\ g^{(m_k-1)}(\lambda_k) = h^{(m_k-1)}(\lambda_k) \quad (k = 1, \ldots, s). \tag{10.39}$$

Ici λ_k $(k = 1, \ldots, s)$ sont les zéros du polynôme minimal $\psi(\lambda)$.

Soit maintenant une certaine fonction $f(\lambda)$. Les nombres

$$f(\lambda_k),\ f'(\lambda_k),\ \ldots,\ f^{(m_k-1)}(\lambda_k)$$
$$(k = 1, \ldots, s;\ m_1 + m_2 + \ldots + m_s = m) \tag{10.40}$$

s'appellent valeurs de la fonction $f(\lambda)$ *dans le spectre de la matrice* A. Désignons l'ensemble de ces valeurs par le symbole $f(\Lambda_A)$.

Si pour la fonction $f(\lambda)$ les valeurs (40) existent, c'est-à-dire si elles ont un sens, on dit que *la fonction* $f(\lambda)$ *est définie dans le spectre de la matrice* A.

Il s'ensuit de la formule (39) que dans le spectre de la matrice A les polynômes $g(\lambda)$ et $h(\lambda)$ ont les mêmes valeurs, c'est-à-dire

$$g(\Lambda_A) = h(\Lambda_A). \tag{10.41}$$

Pour les polynômes $g(\lambda)$ et $h(\lambda)$ la réciproque est également vraie. Si les relations (39) sont données, on en déduit la relation (36) qui implique que $g(A) = h(A)$.

Ainsi, les valeurs du polynôme $g(\lambda)$ dans le spectre de la matrice A déterminent complètement la matrice $g(A)$, c'est-à-dire tous les polynômes $g(\lambda)$ qui dans le spectre de la matrice A prennent les mêmes valeurs ont la même valeur matricielle $g(A)$.

Dans le cas général, la définition de la fonction $f(A)$ est soumise à la même restriction : les valeurs de la fonction $f(\lambda)$ dans le spectre de la matrice A doivent définir complètement $f(A)$, c'est-à-dire toutes les fonctions $f(\lambda)$ prenant les mêmes valeurs dans le spectre de A doivent avoir la même valeur matricielle $f(A)$.

Il en résulte que pour déterminer $f(A)$, il suffit dans le cas général de trouver un polynôme $g(\lambda)$ tel qu'il prenne les mêmes valeurs dans le spectre de la matrice A que la fonction $f(\lambda)$ et de poser $f(A) = g(A)$.

On aboutit ainsi à la définition suivante.

Définition 1. *Si la fonction $f(\lambda)$ est définie dans le spectre de la matrice A, on a*

$$f(A) = g(A), \tag{10.42}$$

où $g(\lambda)$ est un polynôme quelconque qui prend dans le spectre de A les mêmes valeurs que $f(\lambda)$:

$$f(\Lambda_A) = g(\Lambda_A). \tag{10.43}$$

On sait [17, 21] que de tous les polynômes à coefficients complexes qui prennent les mêmes valeurs dans le spectre de A que la fonction $f(\lambda)$ il y a un polynôme $r(\lambda)$ et un seul de degré inférieur ou égal à $m - 1$. Ce polynôme s'appelle *polynôme d'interpolation de Lagrange-Sylvester de la fonction $f(\lambda)$ dans le spectre de la matrice A*. Il est défini univoquement par les conditions d'interpolation

$$r(\lambda_k) = f(\lambda_k),\ r'(\lambda_k) = f'(\lambda_k),\ \ldots,\ r^{(m_k-1)}(\lambda_k) = f^{(m_k-1)}(\lambda_k) \tag{10.44}$$
$$(k = 1, 2, \ldots, s;\ m_1 + m_2 + \ldots + m_s = m).$$

La définition 1 peut être maintenant énoncée sous la forme suivante.

Définition 1'. *Soit $f(\lambda)$ une fonction définie dans le spectre de la matrice A, et $r(\lambda)$, son polynôme d'interpolation de Lagrange-Sylvester défini par les conditions d'interpolation (44). Il vient*

$$f(A) = r(A). \tag{10.45}$$

5. Polynôme d'interpolation de Lagrange-Sylvester. Le polynôme d'interpolation de Lagrange-Sylvester $r(\lambda)$ vérifiant les conditions (44)

$$r(\lambda_k) = f(\lambda_k),\ r'(\lambda_k) = f'(\lambda_k),\ \ldots,\ r^{(m_k-1)}(\lambda_k) = f^{(m_k-1)}(\lambda_k)$$
$$(k = 1, 2, \ldots, s;\ m_1 + m_2 + \ldots + m_s = m),$$

où d'après (29), λ_k ($k = 1, \ldots, s$) sont les zéros du polynôme minimal $\psi(\lambda)$ de la matrice A:

$\psi(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \ldots (\lambda - \lambda_s)^{m_s}$ $(m_1 + m_2 + \ldots + m_s = m)$,
est de la forme

$$r(\lambda) = \sum_{k=1}^{s} \sum_{\rho=0}^{m_k-1} \frac{1}{(m_k-\rho-1)!} \left[\frac{\partial^{m_k-\rho-1}}{\partial\lambda^{m_k-\rho-1}} \frac{f(\lambda)}{\psi_k(\lambda)} \right]_{\lambda=\lambda_k} \frac{\psi(\lambda)}{(\lambda-\lambda_k)^{\rho+1}}, \tag{10.46}$$

où

$$\psi_k(\lambda) = \frac{\psi(\lambda)}{(\lambda-\lambda_k)^{m_k}}. \tag{10.47}$$

Comme il résulte de (46) et (29), le degré du polynôme $r(\lambda)$ est inférieur ou égal à $m - 1$.

Si tous les zéros du polynôme minimal $\psi(\lambda)$ sont simples

$$m_1 = m_2 = \ldots = m_s = 1,$$

conformément à (26) le polynôme d'interpolation $r(\lambda)$ devient

$$r(\lambda) = \sum_{k=1}^{s} \frac{f(\lambda_k)}{\psi_k(\lambda_k)} \frac{\psi(\lambda)}{\lambda-\lambda_k}. \tag{10.48}$$

La formule (48) est la formule d'interpolation de Lagrange bien connue qui, comme on le voit aisément, respecte les conditions $r(\lambda_\mu) = f(\lambda_\mu)$ $(\mu = 1, \ldots, s)$.

On vérifie également sans peine que ce polynôme, défini dans le cas général par l'expression (46), satisfait aux conditions d'interpolation (44). A cet effet, en vertu de (47), il est plus commode de récrire (46) comme suit:

$$r(\lambda) = \sum_{k=1}^{s} \sum_{\rho=0}^{m_k-1} \frac{1}{(m_k-\rho-1)!} \times \\ \times \left[\frac{\partial^{m_k-\rho-1}}{\partial\lambda^{m_k-\rho-1}} \frac{f(\lambda)}{\psi_k(\lambda)} \right]_{\lambda=\lambda_k} (\lambda-\lambda_k)^{m_k-\rho-1} \psi_k(\lambda) \tag{10.49}$$

ou

$$r(\lambda) = \sum_{k=1}^{s} \left\{ \left[\frac{f(\lambda)}{\psi_k(\lambda)} \right]_{\lambda=\lambda_k} + \left[\frac{\partial}{\partial\lambda} \frac{f(\lambda)}{\psi_k(\lambda)} \right]_{\lambda=\lambda_k} (\lambda-\lambda_k) + \right. \\ + \frac{1}{2!} \left[\frac{\partial^2}{\partial\lambda^2} \frac{f(\lambda)}{\psi_k(\lambda)} \right]_{\lambda=\lambda_k} (\lambda-\lambda_k)^2 + \ldots \\ \left. \ldots + \frac{1}{(m_k-1)!} \left[\frac{\partial^{m_k-1}}{\partial\lambda^{m_k-1}} \frac{f(\lambda)}{\psi_k(\lambda)} \right]_{\lambda=\lambda_k} (\lambda-\lambda_k)^{m_k-1} \right\} \psi_k(\lambda). \tag{10.50}$$

Ainsi, par exemple, on trouve d'après (50) que

$$\left(\frac{\partial r}{\partial \lambda}\right)_{\lambda=\lambda_\mu}=\left\{\left[\frac{f(\lambda)}{\psi_\mu(\lambda)}\right]_{\lambda=\lambda_\mu}\psi'_\mu(\lambda)+\left[\frac{\partial}{\partial\lambda}\frac{f(\lambda)}{\psi_\mu(\lambda)}\right]_{\lambda=\lambda_\mu}\psi_\mu(\lambda)\right\}_{\lambda=\lambda_\mu},$$

du fait que les autres termes qui apparaîtront au cours de la dérivation de (49) s'annulent avec $\lambda=\lambda_\mu$. D'où

$$\left(\frac{\partial r}{\partial \lambda}\right)_{\lambda=\lambda_\mu}=\left\{\left[\frac{f(\lambda)}{\psi_\mu(\lambda)}\right]_{\lambda=\lambda_\mu}\psi'_\mu(\lambda)+\right.$$

$$\left.+\left[\frac{f'(\lambda)\,\psi_\mu(\lambda)-f(\lambda)\,\psi'_\mu(\lambda)}{[\psi_\mu(\lambda)]^2}\right]_{\lambda=\lambda_\mu}\psi_\mu(\lambda)\right\}_{\lambda=\lambda_\mu}=f'(\lambda_\mu),$$

ce qui correspond aux conditions (44). D'une façon analogue, on peut justifier les autres conditions (44).

Pour démontrer l'unicité du polynôme de Lagrange-Sylvester, on peut procéder de la façon suivante.

Soit $f(\lambda)$ la fonction scalaire analytique d'une variable complexe λ, régulière dans un domaine qui comprend tous les points λ_k, zéros du polynôme $\psi(\lambda)$ déterminé par l'expression (29)

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\dots(\lambda-\lambda_s)^{m_s}\quad(m_1+m_2+\dots+m_s=m).$$

Dans le domaine considéré, la fonction $\frac{f(\lambda)}{\psi(\lambda)}$ ne possède pas des points singuliers autres que les pôles, et peut s'écrire

$$\frac{f(\lambda)}{\psi(\lambda)}=R(\lambda)+\sum_{k=1}^{s}G_k\left(\frac{1}{\lambda-\lambda_k}\right). \qquad (10.51)$$

Ici $R(\lambda)$ est une fonction régulière dans le domaine considéré; $G_k(1/\lambda-\lambda_k)$, la partie principale du développement de la fonction $f(\lambda)/\psi(\lambda)$ en série de Laurent au voisinage du point $\lambda=\lambda_k$. La fonction $G_k(1/\lambda-\lambda_k)$ peut se mettre sous la forme suivante:

$$G_k\left(\frac{1}{\lambda-\lambda_k}\right)=\sum_{\rho=0}^{m_k-1}\frac{H_{k+\rho}}{(\lambda-\lambda_k)^{\rho+1}}\qquad(k=1,\dots,s). \qquad (10.52)$$

Si l'on porte cette expression dans (51) et multiplie la relation obtenue par $(\lambda-\lambda_k)^{m_k}$, on obtient

$$\frac{f(\lambda)}{\psi_k(\lambda)}=R(\lambda)(\lambda-\lambda_k)^{m_k}+\sum_{k=1}^{s}\sum_{\rho=0}^{m_k-1}H_{k+\rho}(\lambda-\lambda_k)^{m_k-\rho-1}, \qquad (10.53)$$

où $\psi_k(\lambda)$ est déterminée par l'expression (47). Comme la fonction $R(\lambda)(\lambda-\lambda_k)^{m_k}$ est régulière au voisinage du point λ_k et s'annule en ce point avec les dérivées jusqu'à l'ordre (m_k-1), on trouve

à partir de (53) que

$$H_{k+\rho}=\frac{1}{(m_k-\rho-1)!}\left[\frac{\partial^{m_k-\rho-1}}{\partial\lambda^{m_k-\rho-1}}\frac{f(\lambda)}{\psi_k(\lambda)}\right]_{\lambda=\lambda_k}. \tag{10.54}$$

Ainsi, l'expression (53) se met sous la forme

$$\frac{f(\lambda)}{\psi(\lambda)}=R(\lambda)+$$

$$+\sum_{k=1}^{s}\sum_{\rho=0}^{m_k-1}\frac{1}{(m_k-\rho-1)!}\left[\frac{\partial^{m_k-\rho-1}}{\partial\lambda^{m_k-\rho-1}}\frac{f(\lambda)}{\psi_k(\lambda)}\right]_{\lambda=\lambda_k}\frac{1}{(\lambda-\lambda_k)^{\rho+1}}. \tag{10.55}$$

La relation (55) entraîne

$$f(\lambda)=R(\lambda)\,\psi(\lambda)+r(\lambda), \tag{10.56}$$

où $r(\lambda)$ est le polynôme déterminé par l'expression (46). Puisque les fonctions $R(\lambda)$ et $\psi(\lambda)$ sont régulières dans le domaine considéré, et en $\lambda=\lambda_k$ $(k=1,\ldots,s)$ $\psi(\lambda)$ s'annule avec les dérivées d'ordre égal ou inférieur à (m_k-1), $r(\lambda)$ constitue précisément le polynôme qui satisfait aux conditions d'interpolation (44).

Notons que si A est une matrice dont les éléments sont des nombres réels, son polynôme minimal $\psi(\lambda)$, comme on le voit d'après la relation (25), est également à coefficients réels. C'est pourquoi les zéros de $\psi(\lambda)$ sont des nombres réels complexes conjugués deux à deux.

La fonction $f(\lambda)$ est dite *réelle dans le spectre de la matrice A* si pour une valeur propre réelle λ_i toutes ses valeurs dans le spectre $f(\lambda_i), f'(\lambda_i), \ldots$ sont réelles et pour deux valeurs propres complexes conjuguées λ_h et $\lambda_g=\bar{\lambda}_h$ les valeurs associées dans le spectre sont également complexes conjuguées: $f(\lambda_g)=\overline{f(\lambda_h)},\ f'(\lambda_g)=\overline{f'(\lambda_h)},\ldots$ Dans ce cas, les coefficients du polynôme d'interpolation $r(\lambda)$ sont réels, et la matrice $r(A)$ et, par suite, $f(A)=r(A)$, est une matrice à éléments réels.

6. Construction de la fonction e^{At}. D'après (45), la fonction e^{At} s'écrit

$$e^{At}=r(A), \tag{10.57}$$

où $r(\lambda)$ est le polynôme d'interpolation correspondant déterminé par l'expression (46). D'après (46), la formule (57) peut se mettre sous la forme

$$e^{At}=\sum_{k=0}^{m-1}\alpha_k(t)\,A^k, \tag{10.58}$$

où m est le degré du polynôme minimal de la matrice A.

Supposons que la matrice A soit de la forme

$$A=\begin{bmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 2\end{bmatrix}. \tag{10.59}$$

La matrice caractéristique $\lambda E-A$ est alors

$$\lambda E-A=\begin{bmatrix}\lambda-1 & -1 & 0\\ 0 & \lambda-1 & 0\\ 0 & 0 & \lambda-2\end{bmatrix}.$$

Le déterminant de la matrice caractéristique vaut

$$\Delta(\lambda)=(\lambda-1)^2(\lambda-2).$$

La matrice adjointe s'écrit

$$B(\lambda)=\operatorname{adj}(\lambda E-A)=\begin{bmatrix}(\lambda-1)(\lambda-2) & \lambda-2 & 0\\ 0 & (\lambda-1)(\lambda-2) & 0\\ 0 & 0 & (\lambda-1)^2\end{bmatrix}.$$

Le plus grand commun diviseur des éléments de la matrice adjointe

$$D_2(\lambda)=1.$$

D'après (25), le polynôme minimal de la matrice A est

$$\psi(\lambda)=\frac{\Delta(\lambda)}{D_2(\lambda)}=(\lambda-1)^2(\lambda-2). \tag{10.60}$$

En vertu de (40), on a les valeurs suivantes de la fonction $f(\lambda)=e^{\lambda t}$ dans le spectre de la matrice A

$$f(\lambda_1)=(e^{\lambda t})_{\lambda=1}=e^t,\quad f'(\lambda_1)=\left(\frac{\partial}{\partial\lambda}e^{\lambda t}\right)_{\lambda=1}=te^t,$$

$$f(\lambda_2)=(e^{\lambda t})_{\lambda=2}=e^{2t}.$$

Selon (44), les conditions d'interpolation s'écrivent

$$r(\lambda_1)=e^t,\quad \left(\frac{\partial r}{\partial\lambda}\right)_{\lambda=\lambda_1}=te^t,\quad r(\lambda_2)=e^{2t},$$

avec

$$\lambda_1=1,\quad \lambda_2=2.$$

Puisque, d'après (47), pour la matrice A considérée

$$\psi_1(\lambda)=\frac{\psi(\lambda)}{(\lambda-1)^2}=\lambda-2,$$

$$\psi_2(\lambda)=\frac{\psi(\lambda)}{\lambda-2}=(\lambda-1)^2,$$

d'après (50) le polynôme d'interpolation de Lagrange-Sylvester sera

$$r(\lambda)=\left(\frac{e^{\lambda t}}{\lambda-2}\right)_{\lambda=1}(\lambda-2)+\left(\frac{\partial}{\partial\lambda}\frac{e^{\lambda t}}{\lambda-2}\right)_{\lambda=1}(\lambda-1)(\lambda-2)+$$
$$+\left[\frac{e^{\lambda t}}{(\lambda-1)^2}\right]_{\lambda=2}(\lambda-1)^2$$

ou

$$r(\lambda) = -e^t(\lambda - 2) + (-te^t - e^t)(\lambda^2 - 3\lambda + 2) + e^{2t}(\lambda^2 - 2\lambda + 1).$$

Cette expression peut s'écrire de la façon suivante:

$$r(\lambda) = \alpha_0(t) + \alpha_1(t)\lambda + \alpha_2(t)\lambda^2, \tag{10.61}$$

avec

$$\alpha_0(t) = -2te^t + e^{2t},$$
$$\alpha_1(t) = 3te^t + 2e^t - 2e^{2t},$$
$$\alpha_2(t) = -te^t - e^t + e^{2t}.$$

Par conséquent, pour la matrice A considérée (59) et compte tenu du fait que $A^0 = E$ (où E est la matrice unité), on a

$$e^{At} = \alpha_0(t)E + \alpha_1(t)A + \alpha_2(t)A^2. \tag{10.62}$$

Montrons maintenant que les fonctions $\alpha_k(t)$ $(k = 0, 1, \ldots, m - 1)$ qui interviennent dans l'expression (58) sont *linéairement indépendantes*. Considérons d'abord le cas où toutes les valeurs propres de la matrice A sont *simples*. La formule (58) s'écrit alors

$$e^{At} = \sum_{k=0}^{n-1} \alpha_k(t) A^k. \tag{10.63}$$

Dans ce cas, d'après (44), les conditions d'interpolation de la fonction $f(\lambda) = e^{\lambda t}$ définie dans le spectre de la matrice A sont

$$r(\lambda_\rho) = f(\lambda_\rho) = e^{\lambda_\rho t} \qquad (\rho = 1, \ldots, n).$$

Comme d'après (45) la fonction d'une matrice est par définition

$$f(A) = e^{At} = r(A),$$

on obtient, compte tenu de (63), le système d'équations scalaires

$$\sum_{k=0}^{n-1} \lambda_\rho^k \alpha_k(t) = e^{\lambda_\rho t} \qquad (\rho = 1, \ldots, n)$$

ou

$$\alpha_0(t) + \lambda_1\alpha_1(t) + \lambda_1^2\alpha_2(t) + \ldots + \lambda_1^{n-1}\alpha_{n-1}(t) = e^{\lambda_1 t},$$
$$\alpha_0(t) + \lambda_2\alpha_1(t) + \lambda_2^2\alpha_2(t) + \ldots + \lambda_2^{n-1}\alpha_{n-1}(t) = e^{\lambda_2 t},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\alpha_0(t) + \lambda_n\alpha_1(t) + \lambda_n^2\alpha_2(t) + \ldots + \lambda_n^{n-1}\alpha_{n-1}(t) = e^{\lambda_n t}.$$

En désignant par M, $\alpha(t)$ et $R(t)$ les matrices

$$M = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{n-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \lambda_n & \lambda_n^2 & \ldots & \lambda_n^{n-1} \end{bmatrix}, \quad \alpha(t) = \begin{bmatrix} \alpha_0(t) \\ \alpha_1(t) \\ \ldots \\ \alpha_{n-1}(t) \end{bmatrix}, \quad R(t) = \begin{bmatrix} e^{\lambda_1 t} \\ e^{\lambda_2 t} \\ \ldots \\ e^{\lambda_n t} \end{bmatrix},$$

on a l'équation vectorielle suivante

$$M\alpha(t) = R(t).$$

Les éléments du vecteur $R(t)$ sont des fonctions linéairement indépendantes. Le déterminant de la matrice M est un déterminant de Vandermonde et donc $\det M \neq 0$. C'est pourquoi il existe une matrice inverse M^{-1} et, par suite,

$$\alpha(t) = M^{-1}R(t).$$

Puisque $MM^{-1} = E$, $\det M^{-1} = \frac{1}{\det M} \neq 0$, c'est-à-dire les lignes de la matrice M^{-1} sont linéairement indépendantes. C'est pourquoi les éléments du vecteur $\alpha(t)$, c'est-à-dire les *fonctions* $\alpha_0(t)$, $\alpha_1(t)$, . . ., $\alpha_{n-1}(t)$ sont *linéairement indépendantes*.

Une démonstration analogue peut être appliquée au cas où parmi les valeurs propres de la matrice A il y a des valeurs multiples.

Revenons maintenant à notre exemple. Puisque

$$A^2 = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

en vertu de (62) la matrice e^{At} est de la forme

$$e^{At} = \begin{bmatrix} e^t & te^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{2t} \end{bmatrix}. \tag{10.64}$$

Remarquons encore que pour la matrice (59) l'expression (61) du polynôme d'interpolation $r(\lambda)$ peut s'écrire

$$r(\lambda) = e^t\varphi_{11}(\lambda) + te^t\varphi_{12}(\lambda) + e^{2t}\varphi_{21}(\lambda) \tag{10.65}$$

avec

$$\left.\begin{array}{l} \varphi_{11}(\lambda) = -\lambda^2 + 2\lambda, \quad \varphi_{12}(\lambda) = -\lambda^2 + 3\lambda - 2, \\ \varphi_{21}(\lambda) = \lambda^2 - 2\lambda + 1. \end{array}\right\} \tag{10.66}$$

De plus, selon (57) la matrice e^{At} peut se mettre sous la forme

$$e^{At} = e^tZ_{11} + te^tZ_{12} + e^{2t}Z_{21}, \tag{10.67}$$

où

$$Z_{kj} = \varphi_{kj}(A). \tag{10.68}$$

On voit aisément que si au lieu de $e^{\lambda t}$ on prend une autre fonction $f(\lambda)$, alors, conformément à la définition (45) de la fonction d'une matrice, dans le cas de la matrice A déterminée d'après (59) la fonction $f(A)$ peut s'écrire

$$f(A) = f(\lambda_1) Z_{11} + f'(\lambda_1) Z_{12} + f(\lambda_2) Z_{21}. \tag{10.69}$$

Ici $\lambda_1 = 1$, $\lambda_2 = 2$ sont les zéros du polynôme minimal $\psi(\lambda)$ défini par l'expression (60) et dont les ordres de multiplicité sont respectivement $m_1 = 2$, $m_2 = 1$.

7. Composantes de la matrice A. Si dans l'expression (49) du polynôme d'interpolation de Lagrange-Sylvester $r(\lambda)$ on réunit les

termes de même valeur de la fonction $f(\lambda)$ ou d'une quelconque de ses dérivées, le polynôme $r(\lambda)$ devient

$$r(\lambda)=\sum_{k=1}^{s}[f(\lambda_k)\varphi_{k1}(\lambda)+f'(\lambda_k)\varphi_{k2}(\lambda)+\ldots+f^{(m_k-1)}(\lambda_k)\varphi_{km_k}(\lambda)]. \tag{10.70}$$

Ici $\varphi_{kj}(\lambda)$ $(k=1,\ldots,s;\ j=1,\ldots,m_k)$ sont les polynômes en λ à coefficients constants; le degré de ces polynômes est inférieur ou égal à $m-1$, où m est le degré du polynôme minimal $\psi(\lambda)$ de la matrice A.

En vertu de (70) on obtient de la formule (45) l'expression suivante de la fonction $f(A)$:

$$f(A)=\sum_{k=1}^{s}[f(\lambda_k)Z_{k1}+f'(\lambda_k)Z_{k2}+\ldots+f^{(m_k-1)}(\lambda_k)Z_{km_k}], \tag{10.71}$$

où

$$Z_{kj}=\varphi_{kj}(A)\quad(k=1,\ldots,s;\ j=1,\ldots,m_k). \tag{10.72}$$

Les matrices Z_{kj} *ne dépendent pas du choix de la fonction* $f(\lambda)$, elles ne sont déterminées que par la forme de la matrice A. Les valeurs de la fonction $f(\lambda)$ dans le spectre de la matrice A interviennent dans le second membre de l'expression (71). Les matrices Z_{kj} $(k=1,\ldots,s;\ j=1,\ldots,m_k)$ s'appellent *composantes de la matrice A*.

On peut montrer que les composantes Z_{kj} de la matrice A sont *linéairement indépendantes*. En effet, supposons que

$$\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}Z_{kj}=\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}\varphi_{kj}(A)=0. \tag{10.73}$$

Déterminons le polynôme d'interpolation $r(\lambda)$ à partir de m conditions

$$r^{(j-1)}(\lambda_k)=c_{kj}\qquad(k=1,\ldots,s;\ j=1,\ldots,m_k). \tag{10.74}$$

D'après (70) on a

$$r(\lambda)=\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}\varphi_{kj}(\lambda). \tag{10.75}$$

Les expressions (75) et (73) entraînent que

$$r(A)=0. \tag{10.76}$$

Comme le degré du polynôme $r(\lambda)$, déterminé par la formule (75), est inférieur à celui du polynôme minimal $\psi(\lambda)$, la relation (76) conduit à

$$r(\lambda)\equiv 0. \tag{10.77}$$

L'identité (77) et les relations (74) amènent

$$c_{kj} = 0 \qquad (k = 1, \ldots, s; \;\; j = 1, \ldots, m_k). \qquad (10.78)$$

Ainsi la relation (73) ne peut avoir lieu que si les conditions (78) sont observées et, par conséquent, les matrices Z_{kj} linéairement indépendantes.

D'une façon analogue on peut montrer qu'aucune des matrices Z_{kj} *n'est nulle*. En effet, si l'une quelconque des matrices Z_{kj} était nulle, le fait que le degré du polynôme associé $\varphi_{kj}(\lambda)$ est inférieur au degré m du polynôme minimal donnerait lieu à l'identité

$$\varphi_{kj}(\lambda) \equiv 0. \qquad (10.79)$$

Or, ceci est impossible parce que $\varphi_{kj}(\lambda)$ est un polynôme d'interpolation de Lagrange-Sylvester d'une fonction dont toutes les valeurs dans le spectre de la matrice A sont nulles, sauf $f^{(j-1)}(\lambda_k)$ égale à l'unité.

Constatons que dans l'exemple examiné au point 6 les fonctions $\varphi_{kj}(\lambda)$ sont définies par les expressions (66).

8. **Formules générales des composantes Z_{kj} de la matrice A.** Nous avons dit ci-dessus que les matrices Z_{kj} ne sont déterminées que par la forme de la matrice A et ne dépendent pas du choix de la fonction $f(\lambda)$. Indiquons maintenant les formules générales qui définissent les matrices Z_{kj}.

Considérons la fonction

$$f(\mu) = \frac{1}{\lambda - \mu}, \qquad (10.80)$$

où λ est un *paramètre*. On déduit de (80) que

$$\frac{\partial}{\partial\mu} f(\mu) = \frac{1}{(\lambda-\mu)^2}, \quad \frac{\partial^2}{\partial\mu^2} f(\mu) = \frac{2!}{(\lambda-\mu)^3}, \quad \ldots, \quad \frac{\partial^s}{\partial\mu^s} f(\mu) = \frac{s!}{(\lambda-\mu)^{s+1}}, \quad \ldots \qquad (10.81)$$

D'après (80), $f(\mu)(\lambda-\mu) = 1$. Comme $f(\mu)$ et $r(\mu)$, où $r(\mu)$ est le polynôme d'interpolation de Lagrange-Sylvester, coïncident dans le spectre de la matrice A, dans ce même spectre

$$r(\mu)(\lambda - \mu) = 1. \qquad (10.82)$$

En remplaçant dans l'expression (82) l'argument scalaire μ par la matrice A on obtient la relation matricielle:

$$r(A)(\lambda E - A) = E, \qquad (10.83)$$

d'où l'on tire

$$r(A) = (\lambda E - A)^{-1}. \qquad (10.84)$$

En vertu de la relation (45) on en déduit que la fonction $f(\mu)$, donnée par l'expression (80), correspond à

$$f(A) = (\lambda E - A)^{-1}. \qquad (10.85)$$

Les expressions (70), (80) et (81) montrent que le polynôme $r(\mu)$, déterminé par les valeurs $f(\mu)$ dans le spectre de A, est le suivant

$$r(\mu)=\sum_{k=1}^{s}\left[\frac{\varphi_{k1}(\mu)}{\lambda-\lambda_k}+\frac{\varphi_{k2}(\mu)}{(\lambda-\lambda_k)^2}+\frac{2!\,\varphi_{k3}(\mu)}{(\lambda-\lambda_k)^3}+\ldots+\frac{(m_k-1)!\,\varphi_{km_k}(\mu)}{(\lambda-\lambda_k)^{m_k}}\right]. \tag{10.86}$$

D'où, en remplaçant la variable scalaire μ par la matrice A, on obtient

$$f(A)=\sum_{k=1}^{s}\left[\frac{Z_{k1}}{\lambda-\lambda_k}+\frac{Z_{k2}}{(\lambda-\lambda_k)^2}+\frac{2!\,Z_{k3}}{(\lambda-\lambda_k)^3}+\ldots+\frac{(m_k-1)!\,Z_{km_k}}{(\lambda-\lambda_k)^{m_k}}\right]. \tag{10.87}$$

D'après (85), (8) et (19)

$$f(A)=(\lambda E-A)^{-1}=\frac{B(\lambda)}{\Delta(\lambda)}=\frac{C(\lambda)}{\psi(\lambda)}. \tag{10.88}$$

Ici $B(\lambda)$ est la matrice adjointe de la matrice $\lambda E - A$; $\Delta(\lambda)$, le déterminant de la matrice $\lambda E - A$. La matrice $C(\lambda)$ s'appelle matrice adjointe réduite, elle est déterminée par l'expression (14); $\psi(\lambda)$ est le polynôme minimal de A, défini par l'expression (16).

La relation (87) peut maintenant se mettre sous la forme:

$$(\lambda E-A)^{-1}=\frac{C(\lambda)}{\psi(\lambda)}=$$

$$=\sum_{k=1}^{s}\left[\frac{Z_{k1}}{\lambda-\lambda_k}+\frac{Z_{k2}}{(\lambda-\lambda_k)^2}+\frac{2!\,Z_{k3}}{(\lambda-\lambda_k)^3}+\ldots+\frac{(m_k-1)!\,Z_{km_k}}{(\lambda-\lambda_k)^{m_k}}\right]. \tag{10.89}$$

Puisque d'après (29)

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\ldots(\lambda-\lambda_s)^{m_s}$$
$$(m_1+m_2+\ldots+m_s=m),$$

l'expression (89) est la décomposition de la fraction rationnelle $\frac{C(\lambda)}{\psi(\lambda)}$ en une somme de fractions élémentaires, les coefficients de cette décomposition étant les matrices

$$(j-1)!\,Z_{kj} \qquad (k=1,\ldots,s;\; j=1,\ldots,m_k).$$

En comparant les expressions (89) et (55), on trouve que les matrices Z_{kj} sont de la forme

$$Z_{kj}=\frac{1}{(j-1)!\,(m_k-j)!}\left[\frac{\partial^{m_k-j}}{\partial\lambda^{m_k-j}}\frac{C(\lambda)}{\psi_k(\lambda)}\right]_{\lambda=\lambda_k} \tag{10.90}$$

$$(k=1,\ldots,s;\; j=1,\ldots,m_k).$$

9. Représentation des fonctions des matrices par des séries. Adressons-nous de nouveau à la matrice A dont le polynôme minimal $\psi(\lambda)$ est de la forme (29)

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\dots(\lambda-\lambda_s)^{m_s}\quad(m_1+m_2+\dots+m_s=m),$$

et supposons que la fonction $f(\lambda)$ et la suite des fonctions $f_1(\lambda), \dots, f_p(\lambda), \dots$ soient définies dans le spectre de la matrice A.

D é f i n i t i o n 1. *S'il existe des limites*

$$\lim_{p\to\infty} f_p(\lambda_k),\quad \lim_{p\to\infty} f'_p(\lambda_k),\ \dots,\ \lim_{p\to\infty} f_p^{(m_k-1)}(\lambda_k)\qquad (k=1,\dots,s),$$

la suite des fonctions $f_p(\lambda)$ *tend avec* $p\to\infty$ *vers une certaine limite dans le spectre de la matrice* A.

D é f i n i t i o n 2. *La suite des fonctions* $f_p(\lambda)$ *tend avec* $p\to\infty$ *vers la fonction* $f(\lambda)$ *dans le spectre de la matrice* A, *ce qui s'écrit*:

$$\lim_{p\to\infty} f_p(\Lambda_A)=f(\Lambda_A), \tag{10.91}$$

si

$$\lim_{p\to\infty} f_p(\lambda_k)=f(\lambda_k),\quad \lim_{p\to\infty} f'_p(\lambda_k)=f'(\lambda_k),\ \dots$$
$$\dots,\ \lim_{p\to\infty} f_p^{(m_k-1)}(\lambda_k)=f^{(m_k-1)}(\lambda_k)\qquad (k=1,\dots,s). \tag{10.92}$$

La formule (71)

$$f(A)=\sum_{k=1}^{s}[f(\lambda_k)Z_{k1}+f'(\lambda_k)Z_{k2}+\dots+f^{(m_k-1)}(\lambda_k)Z_{km_k}]$$

exprime la fonction $f(A)$ au moyen des valeurs de $f(\lambda)$ dans le spectre de la matrice A. Si l'on considère la matrice comme un vecteur dans l'espace R^{n^2} de dimension n^2, on tire de (71), en vertu de l'indépendance linéaire des matrices Z_{kj}, que toutes les $f(A)$ (pour A donnée) forment un sous-espace de dimension m de R^{n^2} ayant pour base les matrices Z_{kj} ($k=1, \dots, s$; $j=1, \dots, m_k$; $m_1+m_2+\dots+m_s=m$). Dans cette base, les coordonnées du « vecteur » $f(A)$ sont données par m valeurs de la fonction $f(\lambda)$ dans le spectre de la matrice A.

Cette interprétation géométrique rend plus facile la démonstration des théorèmes qui suivent.

T h é o r è m e 1. *Pour que la suite des matrices* $f_p(A)$ *soit convergente pour* $p\to\infty$, *il faut et il suffit que la suite* $f_p(\lambda)$ *soit convergente pour* $p\to\infty$ *dans le spectre de la matrice* A, *c'est-à-dire que les limites* $\lim_{p\to\infty} f_p(A)$, $\lim_{p\to\infty} f_p(\Lambda_A)$ *existent toujours simultanément.*

Dans ce cas l'égalité

$$\lim_{p\to\infty} f_p(\Lambda_A) = f(\Lambda_A) \qquad (10.93)$$

entraîne l'égalité

$$\lim_{p\to\infty} f_p(A) = f(A) \qquad (10.94)$$

et inversement.

Démonstration. Si la suite $f_p(\lambda)$ est convergente pour $p \to \infty$ dans le spectre de la matrice A, la formule

$$f_p(A) = \sum_{k=1}^{s} [f_p(\lambda_k) Z_{k1} + f'_p(\lambda_k) Z_{k2} + \ldots + f_p^{(m_k-1)}(\lambda_k) Z_{km_k}] \qquad (10.95)$$

entraîne l'existence de la limite $\lim\limits_{p\to\infty} f_p(A)$. Les formules (95) et (71) permettent de conclure que si la relation (93) a lieu, la relation (94) est également vraie.

Passons à la démonstration de la réciproque. Supposons qu'il existe une $\lim\limits_{p\to\infty} f_p(A)$ et que le type de la matrice $f_p(A)$ est $n \times n$, c'est-à-dire qu'elle possède n^2 éléments; de plus, $n \geqslant m$, puisque m est le degré du polynôme minimal de la matrice A. La relation matricielle (95) est équivalente à n^2 relations scalaires. Choisissons-en m relations que nous allons considérer comme un système de m équations algébriques linéaires par rapport aux inconnues

$$f_p(\lambda_k),\ f'_p(\lambda_k),\ \ldots,\ f_p^{(m_k-1)}(\lambda_k)$$
$$(k = 1, \ldots, s;\ m_1 + m_2 + \ldots + m_s = m).$$

Les matrices Z_{kj} étant linéairement indépendantes, parmi les n^2 relations il y aura m relations telles dont le déterminant composé de coefficients affectés à $f_p(\lambda_k)$, $f'_p(\lambda_k)$, ..., $f_p^{(m_k-1)}(\lambda_k)$ sera non nul. Ainsi, les valeurs $f_p(\lambda_k), f'_p(\lambda_k), \ldots, f_p^{(m_k-1)}(\lambda_k)$ $(k = 1, \ldots, s)$ peuvent être exprimées à l'aide des formes linéaires de m éléments de la matrice $f_p(A)$. Il en résulte l'existence de la limite $\lim\limits_{p\to\infty} f_p(\Lambda_A)$ et par suite, si la relation (94) a lieu, la relation (93) est également vérifiée.

Il ressort du théorème démontré que si la suite des polynômes $g_p(\lambda)$ $(p = 1, 2, \ldots)$ tend vers la fonction $f(\lambda)$ dans le spectre de la matrice A,

$$\lim_{p\to\infty} g_p(A) = f(A).$$

Définition 3. *La série* $\sum\limits_{p=0}^{\infty} u_p(\lambda)$ *converge dans le spectre de la matrice* A *vers la fonction* $f(\lambda)$, *ce qui se note*

$$f(\Lambda_A) = \sum_{p=0}^{\infty} u_p(\Lambda_A), \qquad (10.96)$$

si toutes les fonctions qui figurent ici sont définies dans le spectre de la matrice A et donnent lieu aux égalités

$$f(\lambda_k)=\sum_{p=0}^{\infty} u_p(\lambda_k),\quad f'(\lambda_k)=\sum_{p=0}^{\infty} u'_p(\lambda_k),\ \ldots,\ f^{(m_k-1)}(\lambda_k)=$$

$$=\sum_{p=0}^{\infty} u_p^{(m_k-1)}(\lambda_k)\qquad (k=1,\ \ldots,\ s),$$

les seconds membres de ces égalités étant constitués de séries convergentes.

Si l'on introduit les notations

$$s_p(\lambda)=\sum_{q=0}^{p} u_q(\lambda)\qquad (p=0,\ 1,\ 2,\ \ldots),\tag{10.97}$$

la relation (96) peut s'écrire

$$f(\Lambda_A)=\lim_{p\to\infty} s_p(\Lambda_A).\tag{10.98}$$

Maintenant on peut donner au théorème 1, démontré dans ce qui précède, une autre formulation équivalente.

Théorème 1'. *Pour que la série* $\sum_{p=0}^{\infty} u_p(A)$ *converge vers une certaine matrice, il faut et il suffit que la série* $\sum_{p=0}^{\infty} u_p(\lambda)$ *converge dans le spectre de la matrice A. L'égalité*

$$f(\Lambda_A)=\sum_{p=0}^{\infty} u_p(\Lambda_A)$$

entraîne alors l'égalité

$$f(A)=\sum_{p=0}^{\infty} u_p(A)$$

et inversement.

Considérons maintenant une *série entière* à cercle de convergence $|\lambda-\lambda_0|<R$ et à somme $f(\lambda)$

$$f(\lambda)=\sum_{p=0}^{\infty} \alpha_p(\lambda-\lambda_0)^p\qquad (|\lambda-\lambda_0|<R).\tag{10.99}$$

Comme une série entière est dérivable terme à terme n'importe quel nombre de fois à l'intérieur du cercle de convergence, la série (99) converge dans le spectre de n'importe quelle matrice dont les valeurs propres se situent à l'intérieur de ce cercle. On en déduit le théorème suivant.

Théorème 2. *Si la fonction $f(\lambda)$ se développe en une série entière dans le cercle $|\lambda - \lambda_0| < R$:*

$$f(\lambda) = \sum_{p=0}^{\infty} \alpha_p (\lambda - \lambda_0)^p \qquad (10.100)$$

ce développement garde son sens si la variable scalaire λ est remplacée par une matrice A quelconque dont les valeurs propres reposent à l'intérieur du cercle de convergence.

Ce théoreme entraîne, par exemple, les développements suivants :

$$\left.\begin{array}{ll} e^A = \sum\limits_{p=0}^{\infty} \dfrac{A^p}{p!}, & \cos A = \sum\limits_{p=0}^{\infty} \dfrac{(-1)^p}{(2p)!} A^{2p}, \\ & \sin A = \sum\limits_{p=0}^{\infty} (-1)^p \dfrac{A^{2p+1}}{(2p+1)!}, \\ \operatorname{ch} A = \sum\limits_{p=0}^{\infty} \dfrac{A^{2p}}{(2p)!}, & \operatorname{sh} A = \sum\limits_{p=0}^{\infty} \dfrac{A^{2p+1}}{(2p+1)!}, \end{array}\right\} \qquad (10.101)$$

$$(E - A)^{-1} = \sum_{p=0}^{\infty} A^p \qquad (|\lambda_k| < 1 ;\ k = 1, \ldots, s). \qquad (10.102)$$

Les développements (101) ont lieu pour toute matrice A du fait que les développements des fonctions correspondantes par rapport à la variable scalaire λ ont lieu pour $|\lambda| < \infty$.

10. Extension de la formule intégrale de Cauchy pour les fonctions analytiques aux fonctions des matrices. Considérons dans le plan de la variable complexe λ le domaine limité par le contour fermé Γ et contenant les valeurs propres $\lambda_1, \ldots, \lambda_s$ de la matrice A. Soit $f(\lambda)$ la fonction analytique arbitraire régulière dans ce domaine y compris la frontière Γ. D'après les formules de Cauchy

$$f(\lambda_k) = \frac{1}{2\pi i} \int_\Gamma \frac{f(\lambda)}{\lambda - \lambda_k} d\lambda, \quad f'(\lambda_k) = \frac{1}{2\pi i} \int_\Gamma \frac{f(\lambda)}{(\lambda - \lambda_k)^2} d\lambda, \ \ldots, \ f^{(m_k - 1)}(\lambda_k) =$$

$$= \frac{(m_k - 1)!}{2\pi i} \int_\Gamma \frac{f(\lambda)}{(\lambda - \lambda_k)^{m_k}} d\lambda \qquad (k = 1, \ldots, s). \qquad (10.103)$$

En multipliant les deux membres de la relation matricielle (89) par $\frac{f(\lambda)}{2\pi i}$ et en intégrant le long du contour fermé Γ, on obtient

$$\frac{1}{2\pi i} \int_\Gamma (\lambda E - A)^{-1} f(\lambda)\, d\lambda =$$

$$= \sum_{k=1}^{s} [f(\lambda_k) Z_{k1} + f'(\lambda_k) Z_{k2} + \ldots + f^{(m_k - 1)}(\lambda_k) Z_{km_k}]. \qquad (10.104)$$

En tenant compte de la formule (71), la relation (104) se met sous la forme

$$f(A)=\frac{1}{2\pi i}\int_{\Gamma}(\lambda E-A)^{-1}f(\lambda)\,d\lambda. \qquad (10.105)$$

La formule (105) peut être adoptée comme définition de la *fonction analytique d'une matrice*.

11. Certaines propriétés des fonctions des matrices.

Théorème 1. *Soient* $G(u_1, u_2, \ldots, u_l)$ *le polynôme en* $u_1, u_2, \ldots, u_l$; $f_1(\lambda), f_2(\lambda), \ldots, f_l(\lambda)$, *les fonctions de* λ *définies dans le spectre de la matrice* A *et*

$$g(\lambda) = G[f_1(\lambda), f_2(\lambda), \ldots, f_l(\lambda)]. \qquad (10.106)$$

Alors, si dans le spectre de A *la fonction* $g(\lambda)$ *s'annule*

$$g(\Lambda_A) = 0, \qquad (10.107)$$

on a la relation suivante:

$$G[f_1(A), f_2(A), \ldots, f_l(A)] = 0. \qquad (10.108)$$

Démonstration. Désignons par $r_1(\lambda), r_2(\lambda), \ldots, r_l(\lambda)$ les polynômes d'interpolation de Lagrange-Sylvester pour $f_1(\lambda)$, $f_2(\lambda), \ldots, f_l(\lambda)$ et supposons que

$$h(\lambda) = G[r_1(\lambda), r_2(\lambda), \ldots, r_l(\lambda)]. \qquad (10.109)$$

Puisque dans le spectre de la matrice A les valeurs de $g(\lambda)$ et $h(\lambda)$ coïncident, la relation (107) entraîne que

$$h(\Lambda_A) = 0. \qquad (10.110)$$

Mais alors la formule (71) conduit à

$$h(A) = 0 \qquad (10.111)$$

ou

$$G[r_1(A), r_2(A), \ldots, r_l(A)]=G[f_1(A), f_2(A), \ldots, f_l(A)]=0, \qquad (10.112)$$

ce qu'il fallait démontrer.

Examinons à titre de corollaire du théorème 1 les exemples suivants.

1°. Soit

$$G(u_1, u_2) = u_1^2 + u_2^2 - 1,$$
$$f_1(\lambda) = \sin\lambda, \; f_2(\lambda) = \cos\lambda.$$

Il vient

$$g(\lambda) = G[f_1(\lambda), f_2(\lambda)] = \sin^2\lambda + \cos^2\lambda - 1.$$

Etant donné que le spectre de toute matrice A donne lieu pour la fonction considérée à la relation,

$$g(\Lambda_A) = 0$$

on a, d'après (112),

$$G[f_1(A), f_2(A)] = \sin^2 A + \cos^2 A - E = 0$$

ou

$$\sin^2 A + \cos^2 A = E. \qquad (10.113)$$

2°. Soit

$$G(u_1, u_2) = u_1 u_2 - 1,$$

$$f_1(\lambda) = e^{\lambda}, \quad f_2(\lambda) = e^{-\lambda}.$$

Il vient

$$g(\lambda) = G[f_1(\lambda), f_2(\lambda)] = e^{\lambda} e^{-\lambda} - 1.$$

Dans le spectre de toute matrice A on a pour la fonction considérée $g(\lambda)$

$$g(\Lambda_A) = 0.$$

C'est pourquoi, en vertu de (112),

$$G[f_1(A), f_2(A)] = e^A e^{-A} - E = 0$$

ou

$$e^A e^{-A} = E. \qquad (10.114)$$

Il ressort de la relation (114) que la matrice inverse $(e^A)^{-1}$ est de la forme

$$(e^A)^{-1} = e^{-A}. \qquad (10.115)$$

3°. Soit

$$G(u_1, u_2, u_3) = u_1 u_2 - u_3,$$

$$f_1(\lambda) = e^{\lambda t_1}, \quad f_2(\lambda) = e^{\lambda t_2}, \quad f_3(\lambda) = e^{\lambda(t_1+t_2)}.$$

Alors,

$$g(\lambda) = G[f_1(\lambda), f_2(\lambda), f_3(\lambda)] = e^{\lambda t_1} e^{\lambda t_2} - e^{\lambda(t_1+t_2)}.$$

Dans le spectre de toute matrice A on a pour la fonction considérée $g(\lambda)$ la relation

$$g(\Lambda_A) = 0.$$

D'après (112) on en tire

$$e^{At_1} e^{At_2} - e^{A(t_1+t_2)} = 0$$

ou

$$e^{At_1} e^{At_2} = e^{A(t_1+t_2)}. \qquad (10.116)$$

4°. Soit

$$G(u_1, u_2, u_3) = u_1 + iu_2 - u_3,$$

$$f_1(\lambda) = \cos\lambda, \quad f_2(\lambda) = \sin\lambda, \quad f_3(\lambda) = e^{i\lambda}.$$

Il vient

$$g(\lambda) = G[f_1(\lambda), f_2(\lambda), f_3(\lambda)] = \cos\lambda + i\sin\lambda - e^{i\lambda}.$$

Dans le spectre de toute matrice A on a pour la fonction $g(\lambda)$ la relation

$$g(\Lambda_A) = 0.$$

Il en résulte d'après (112)

$$\cos A + i \sin A - e^{iA} = 0$$

ou

$$e^{iA} = \cos A + i \sin A. \qquad (10.117)$$

Ainsi le théorème 1 donne les conditions (107) sous lesquelles les identités associant les fonctions de la variable scalaire λ peuvent être étendues aux valeurs matricielles de la variable.

Le théorème 1 peut être renforcé pour démontrer le théorème suivant ([21], p. 121).

Théorème 2. *Soit*

$$g(\lambda) \equiv G[f_1(\lambda), f_2(\lambda), \ldots, f_l(\lambda)],$$

où les fonctions $f_1(\lambda), f_2(\lambda), \ldots, f_l(\lambda)$ *sont définies dans le spectre de la matrice* A, *alors que la fonction* $G(u_1, u_2, \ldots, u_l)$ *résulte de l'application successive aux variables* $u_1, u_2, \ldots, u_l$ *des opérations d'addition, de multiplication, de multiplication par un nombre et de la substitution de la valeur variable par une fonction arbitraire de celle-ci. Alors, si dans le spectre de* A *la fonction* $g(\lambda)$ *s'annule*

$$g(\Lambda_A) = 0,$$

on a la relation suivante:

$$G[f_1(A), f_2(A), \ldots, f_l(A)] = 0.$$

12. Intégration des systèmes d'équations différentielles linéaires à coefficients constants à l'aide des fonctions des matrices. Cherchons la solution du système d'équations différentielles linéaires

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} A_{jk} x_k \qquad (j = 1, \ldots, n) \qquad (10.118)$$

sous les conditions initiales

$$x_j = x_j(0) \qquad (j = 1, \ldots, n) \quad \text{pour} \quad t = 0. \qquad (10.119)$$

Ici A_{jk} $(j, k = 1, \ldots, n)$ sont des constantes, en général complexes.

Introduisons les matrices

$$A = \begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ A_{n1} & \ldots & A_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \ldots \\ x_n \end{bmatrix}. \qquad (10.120)$$

La dérivation (intégration) des matrices consistant à dériver (intégrer) chacun de leurs éléments, le système d'équations différentielles

scalaires (118) peut être remplacé par une équation différentielle vectorielle

$$\dot{x} = Ax. \tag{10.121}$$

Les conditions initiales (119) peuvent s'écrire

$$x = x(0) \quad \text{pour} \quad t = 0. \tag{10.122}$$

Mettons le vecteur cherché $x(t)$ sous la forme d'une série de MacLaurin par rapport aux puissances de t:

$$x(t) = x(0) + \dot{x}(0)\,t + \ddot{x}(0)\frac{t^2}{2!} + \ldots + x^{(n)}(0)\frac{t^n}{n!} + \ldots \tag{10.123}$$

En dérivant le premier et le second membre de l'équation (121), on tombe sur

$$\ddot{x} = A\dot{x} = A^2x, \quad \dddot{x} = A^2\dot{x} = A^3x, \quad \ldots, \quad x^{(n)} = A^nx, \quad \ldots \tag{10.124}$$

Les relations (121) et (124) ont lieu pour toute valeur de la variable t. Avec $t = 0$ elles deviennent

$$\dot{x}(0) = Ax(0), \quad \ddot{x}(0) = A^2x(0), \quad \ldots, \quad x^{(n)}(0) = A^nx(0), \quad \ldots \tag{10.125}$$

La série (123) se met ainsi sous la forme

$$x(t) = x(0) + tAx(0) + \frac{t^2}{2!}A^2x(0) + \ldots + \frac{t^n}{n!}A^nx(0)\ldots =$$
$$= \left(E + At + \frac{A^2t^2}{2!} + \ldots + \frac{A^nt^n}{n!} + \ldots\right)x(0), \tag{10.126}$$

c'est-à-dire selon (101)

$$x(t) = e^{At}x(0). \tag{10.127}$$

Comme

$$\frac{d}{dt}e^{At} = \frac{d}{dt}\left(E + At + \frac{A^2t^2}{2!} + \ldots\right) = A + A^2t + \frac{A^3t^2}{2!} + \ldots = Ae^{At}, \tag{10.128}$$

la substitution immédiate dans l'équation différentielle (121) montre que l'expression (127) est une solution de cette équation, qui vérifie les conditions initiales (122).

Puisque (127) entraîne que

$$x(t_0) = e^{At_0}x(0),$$

et d'après (116)

$$e^{At}x(0) = e^{A(t-t_0)}e^{At_0}x(0) = e^{A(t-t_0)}x(t_0),$$

l'expression (127) peut être ramenée à la forme

$$x(t) = e^{A(t-t_0)}x(t_0). \tag{10.129}$$

L'expression (129) est la solution de l'équation différentielle vectorielle (121) qui satisfait aux conditions initiales

$$x = x(t_0) \quad \text{pour} \quad t = t_0. \tag{10.130}$$

Considérons maintenant l'expression (71). En posant

$$f(\lambda) = e^{\lambda t}, \tag{10.131}$$

on a

$$f(\lambda_k) = e^{\lambda_k t}, \quad f'(\lambda_k) = \left(\frac{\partial}{\partial\lambda} e^{\lambda t}\right)_{\lambda=\lambda_k} = te^{\lambda_k t}, \quad \ldots, \quad f^{(m_k-1)}(\lambda_k) =$$

$$= \left(\frac{\partial^{m_k-1}}{\partial\lambda^{m_k-1}} e^{\lambda t}\right)_{\lambda=\lambda_k} = t^{m_k-1} e^{\lambda_k t}. \tag{10.132}$$

En vertu de (71), (131) et (132), la fonction e^{At} peut s'écrire

$$e^{At} = \sum_{k=1}^{s} (Z_{k1} + Z_{k2}t + \ldots + Z_{km_k} t^{m_k-1}) e^{\lambda_k t}, \tag{10.133}$$

et la solution de (127) peut être ramenée à une forme contenant les composantes de la matrice A.

Si l'on tient compte de la relation (133) on voit sans peine que la solution du système d'équations différentielles (118) obtenue en appliquant les fonctions d'une matrice coïncide (comme il se doit) avec la solution (4.124) obtenue au § 4. (Au § 4 le système examiné est du type plus général.) Cette correspondance justifie dans une certaine mesure la définition de la fonction d'une matrice adoptée dans la théorie des matrices (45).

Examinons maintenant le système d'équations linéaires inhomogènes

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} A_{jk} x_k + f_j(t) \qquad (j = 1, \ldots, n). \tag{10.134}$$

Le système d'équations différentielles scalaires (134) peut être remplacé par une équation différentielle vectorielle

$$\dot{x} = Ax + f(t), \tag{10.135}$$

où par $f(t)$ on désigne le vecteur

$$f(t) = \begin{bmatrix} f_1(t) \\ \cdots \\ f_n(t) \end{bmatrix}.$$

Cherchons la solution de l'équation (135) qui vérifie les conditions initiales (130) sous la forme

$$x(t) = e^{At} z(t), \tag{10.136}$$

où

$$z(t)=\begin{bmatrix} z_1(t) \\ \cdots \\ z_n(t) \end{bmatrix}. \tag{10.137}$$

Les éléments du vecteur $z(t)$ sont à déterminer.

En portant l'expression (136) dans (135), on obtient

$$e^{At}\dot{z}=f(t)$$

ou

$$\dot{z}=e^{-At}f(t). \tag{10.138}$$

On en tire

$$z(t)=C+\int_{t_0}^{t} e^{-A\tau}f(\tau)\,d\tau, \tag{10.139}$$

avec C le vecteur composé de constantes arbitraires

$$C=\begin{bmatrix} C_1 \\ \cdots \\ C_n \end{bmatrix}. \tag{10.140}$$

Maintenant l'expression (136) se met sous la forme

$$x(t)=e^{At}\left(C+\int_{t_0}^{t} e^{-A\tau}f(\tau)\,d\tau\right). \tag{10.141}$$

L'expression (141) entraîne que

$$[x(t)]_{t=t_0}=e^{At_0}C,$$

d'où, compte tenu des conditions initiales (130),

$$C=e^{-At_0}x(t_0). \tag{10.142}$$

En portant dans (141) la valeur trouvée de C, on obtient

$$x(t)=e^{A(t-t_0)}x(t_0)+\int_{t_0}^{t} e^{A(t-\tau)}f(\tau)\,d\tau. \tag{10.143}$$

C'est précisément la solution de l'équation différentielle vectorielle (135) sous les conditions initiales (130).

13. Comparaison avec les solutions fournies par la transformation de Laplace. Montrons maintenant l'identité de la solution (143) et de la solution qu'on peut obtenir par les méthodes exposées au § 4, fondées sur l'application de la transformation de Laplace.

En introduisant les notations

$$\xi(p) \div x(t), \quad \widetilde{f}(p) \div f(t), \tag{10.144}$$

on obtient qu'à l'équation différentielle vectorielle (135)

$$\dot{x} = Ax + f(t)$$

correspond la transformation suivante

$$(pE - A)\,\xi(p) = px(0) + \widetilde{f}(p), \tag{10.145}$$

où $x(0)$ est la valeur de $x(t)$ avec $t = 0$. D'où

$$\xi(p) = p(pE - A)^{-1}x(0) + (pE - A)^{-1}\widetilde{f}(p). \tag{10.146}$$

On a en vertu de (89)

$$p(pE-A)^{-1} = \sum_{k=1}^{s}\left[Z_{k1}\frac{p}{p-\lambda_k} + Z_{k2}\frac{p}{(p-\lambda_k)^2} + \right.$$
$$\left. + Z_{k3}\frac{2!\,p}{(p-\lambda_k)^3} + \ldots + Z_{km_k}\frac{(m_k-1)!\,p}{(p-\lambda_k)^{m_k}}\right]. \tag{10.147}$$

Désignons par $N(t)$ l'original dont la transformée est la fonction $p(pE - A)^{-1}$

$$p(pE - A)^{-1} \div N(t). \tag{10.148}$$

Compte tenu du fait que

$$\frac{p}{(p-\lambda_k)^{\mu+1}} \div \frac{t^{\mu}e^{\lambda_k t}}{\mu!}, \tag{10.149}$$

on trouve

$$N(t) = \sum_{k=1}^{s}\left[Z_{k1} + Z_{k2}t + \ldots + Z_{km_k}t^{m_k-1}\right]e^{\lambda_k t}. \tag{10.150}$$

En vertu du théorème de multiplication des transformées

$$\frac{1}{p}\,p(pE-A)^{-1}\widetilde{f}(p) \div \int_0^t N(t-\tau)\,f(\tau)\,d\tau. \tag{10.151}$$

Ainsi à la transformée (146) correspond l'original

$$x(t) = N(t)\,x(0) + \int_0^t N(t-\tau)\,f(\tau)\,d\tau. \tag{10.152}$$

La comparaison des expressions (150) et (133) montre que

$$N(t) = e^{At}, \tag{10.153}$$

ainsi (152) se met sous la forme

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}f(\tau)\,d\tau. \qquad (10.154)$$

D'après (154)

$$x(t_0) = e^{At_0}x(0) + \int_0^{t_0} e^{A(t_0-\tau)}f(\tau)\,d\tau. \qquad (10.155)$$

L'expression (154) peut être ramenée à la forme

$$x(t) = e^{A(t-t_0)}\left[e^{At_0}x(0) + \int_0^{t_0} e^{A(t_0-\tau)}f(\tau)\,d\tau\right] + \int_{t_0}^t e^{A(t-\tau)}f(\tau)\,d\tau, \qquad (10.156)$$

d'où, en tenant compte de (155), on obtient

$$x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^t e^{A(t-\tau)}f(\tau)\,d\tau, \qquad (10.157)$$

ce qui coïncide avec l'expression (143).

En remplaçant conformément à (153) la fonction e^{At} par $N(t)$, récrivons l'expression (157) sous la forme:

$$x(t) = N(t-t_0)\,x(t_0) + \int_{t_0}^t N(t-\tau)\,f(\tau)\,d\tau. \qquad (10.158)$$

Remarquons encore qu'en vertu de (153), (114) et (115), la matrice $N(t)$ définie par l'expression (150) et la matrice inverse $N^{-1}(t)$ vérifient les relations suivantes:

$$N(0) = E, \qquad (10.159)$$

$$N^{-1}(t) = N(-t), \qquad (10.160)$$

$$N(t)\,N^{-1}(\tau) = N(t-\tau). \qquad (10.161)$$

Compte tenu du fait que

$$e^{At} = E + At + \frac{A^2t^2}{2!} + \dots$$

et que lors de la multiplication, les puissances des matrices sont commutatives, on obtient encore

$$AN = NA. \qquad (10.162)$$

Les relations (153) et (128) entraînent également

$$\frac{dN}{dt} = AN. \qquad (10.163)$$

§ 11. Commandabilité et observabilité des systèmes linéaires

1. Commandabilité des systèmes décrits par des équations différentielles linéaires à coefficients constants. Considérons un système décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = Ax + Gu, \tag{11.1}$$

où

$$x = \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \dots & \dots & \dots \\ A_{n1} & \dots & A_{nn} \end{bmatrix},$$

$$G = \begin{bmatrix} G_{11} & \dots & G_{1r} \\ \dots & \dots & \dots \\ G_{n1} & \dots & G_{nr} \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \dots \\ u_r \end{bmatrix}. \tag{11.2}$$

Nous avons désigné ici par x_k $(k = 1, \ldots, n)$ les variables qui déterminent l'état du système; par u_l $(l = 1, \ldots, r)$, les *commandes* appliquées au système. On suppose que les éléments des matrices A et G sont constants.

L'équation différentielle vectorielle (1) est équivalente au système d'équations différentielles scalaires

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} A_{jk}x_k + \sum_{l=1}^{r} G_{jl}u_l \qquad (j = 1, \ldots, n). \tag{11.3}$$

Les équations (3) sont des équations différentielles linéaires à coefficients constants ; on peut les interpréter, par exemple, comme des équations aux variations (ou linéarisées) par rapport à un mouvement stationnaire (ou par rapport à l'état d'équilibre) d'un système de points matériels. Soit s le nombre de degrés de liberté de ce système, et $q_1, \ldots, q_s$, ses coordonnées généralisées. Le rang du système d'équations (3) sera alors

$$n = 2s.$$

Les variables x_i $(i = 1, \ldots, n)$ des équations (3) peuvent être les coordonnées de phase ou variables canoniques (coordonnées généralisées et impulsions généralisées) du système considéré de points matériels ou leur être associées par une certaine transformation linéaire.

Puisque le nombre de forces généralisées appliquées à un système de points matériels ne peut dépasser le nombre de degrés de liberté, le nombre r des commandes u_l $(l = 1, \ldots, r)$ des équations différentielles (3) doit vérifier la condition

$$r \leqslant s.$$

Notons encore que celles des équations du système (3) qui n'expriment que la relation entre les variables x_j, par exemple les équations du type

$$\dot{x}_\mu = x_{\mu+1}, \tag{11.4}$$

exigent, dans le cas où x_j sont des coordonnées de phase du système, qu'on ait les identités

$$\sum_{l=1}^{r} G_{\mu l} u_l \equiv 0 \qquad (\mu = \mu_1, \ldots, \mu_s), \tag{11.5}$$

ce qui exige que les lignes correspondantes de la matrice G soient composées d'éléments nuls

$$G_{\mu 1} = G_{\mu 2} = \ldots = G_{\mu r} = 0 \qquad (\mu = \mu_1, \ldots, \mu_s). \tag{11.6}$$

Toutefois dans le cas général, la structure d'un système de commande est plus complexe. Il comporte encore des organes de production des signaux de commande, etc. En outre, les équations (3) peuvent compter également des équations différentielles qui décrivent les programmes des calculateurs de commande appartenant au système, etc. Ces équations peuvent contenir également des commandes à elles. Dans ce cas on aura la relation $n > 2s$, où par s on entend le nombre de degrés de liberté seulement de l'ensemble des éléments mécaniques appartenant au système (3) ; il se peut également que $r > s$.

Examinons maintenant le problème qui consiste à ramener le système décrit par l'équation (1) d'un état initial donné à un état imposé quelconque pendant un intervalle de temps fini, en choisissant d'une façon convenable la loi de variation régissant les commandes $u_l = u_l(t)$ $(l = 1, \ldots, r)$.

La propriété qui vient d'être formulée s'appelle *commandabilité*. Les systèmes qui en jouissent sont dits *complètement commandables*.

La commandabilité d'un système étant définie par la structure des matrices A et G, cette notion est également associée à ces matrices et l'on dit que le couple (A, G) est commandable ou, respectivement, non commandable.

Passons maintenant à la résolution du problème posé. D'après (10.143), la loi du mouvement du système décrit par l'équation (1) est de la forme :

$$x(t) = e^{At} x(0) + \int_0^t e^{A(t-\tau)} G u(\tau)\, d\tau. \tag{11.7}$$

Supposons qu'il existe une loi de variation des commandes $u_l = u_l(t)$ $(l = 1, \ldots, r)$ telle qu'elle permet de ramener le

système à l'instant $t=T$ à l'origine des coordonnées, c'est-à-dire assure l'observation de la condition

$$x(T)=0. \tag{11.8}$$

Puisque d'après (7)

$$x(T)=e^{AT}x(0)+\int_0^T e^{A(T-\tau)}Gu(\tau)\,d\tau, \tag{11.9}$$

on a en vertu de (8) la relation

$$\int_0^T e^{A(T-\tau)}Gu(\tau)\,d\tau=-e^{AT}x(0). \tag{11.10}$$

En multipliant le premier et le deuxième membre de la relation (10) par e^{-AT}, on obtient

$$\int_0^T e^{-A\tau}Gu(\tau)\,d\tau=-x(0). \tag{11.11}$$

D'après (10.58)

$$e^{-A\tau}=\sum_{k=0}^{m-1}\alpha_k(-\tau)A^k, \tag{11.12}$$

où m est le degré du polynôme minimal $\psi(\lambda)$ de la matrice A, et $\alpha_k(t)$ les coefficients du polynôme d'interpolation de Lagrange-Sylvester $r(\lambda)$ construit pour la fonction $e^{\lambda t}$ définie dans le spectre de la matrice A.

En tenant compte de l'expression (12), la relation (11) peut s'écrire

$$\sum_{k=0}^{m-1}A^kG\int_0^T u(\tau)\alpha_k(-\tau)\,d\tau=-x(0) \tag{11.13}$$

ou

$$G\int_0^T u(\tau)\alpha_0(-\tau)\,d\tau+AG\int_0^T u(\tau)\alpha_1(-\tau)\,d\tau+\ldots$$
$$\ldots+A^{m-1}G\int_0^T u(\tau)\alpha_{m-1}(-\tau)\,d\tau=-x(0). \tag{11.14}$$

Chacun des termes du premier membre de la relation (14) est un vecteur d'ordre $n\times 1$.

Le premier membre de la relation (14) peut être mis sous la forme d'un produit de deux matrices; la relation (14) s'écrit alors

$$[G\,AG\,A^2G\,\ldots\,A^{m-1}G]\times\begin{bmatrix}\int_0^T u(\tau)\,\alpha_0(-\tau)\,d\tau\\ \int_0^T u(\tau)\,\alpha_1(-\tau)\,d\tau\\ \cdots\cdots\cdots\cdots\\ \int_0^T u(\tau)\,\alpha_{m-1}(-\tau)\,d\tau\end{bmatrix}=-x(0).\qquad(11.15)$$

Comme A^kG ($k=0, 1, \ldots, m-1$) est une matrice $n\times r$,

$$W=[G\;AG\;A^2G\;\ldots\;A^{m-1}G]\qquad(11.16)$$

est une matrice partitionnée $1\times m$.

Faisons la remarque suivante. Selon (10.29), le polynôme minimal de la matrice A s'écrit

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\ldots(\lambda-\lambda_s)^{m_s}\quad(m_1+m_2+\ldots+m_s=m)$$

ou

$$\psi(\lambda)=\lambda^m+c_1\lambda^{m-1}+\ldots+c_{m-1}\lambda+c_m.$$

Etant donné que $\psi(\lambda)$ est un polynôme annulateur de la matrice A, c'est-à-dire

$$\psi(A)=0,$$

on a

$$A^m+c_1A^{m-1}+\ldots+c_{m-1}A+c_mE=0.$$

Il en résulte que

$$A^m=-c_1A^{m-1}-\ldots-c_{m-1}A-c_mE,$$

c'est-à-dire que la matrice A^m est une combinaison linéaire des matrices E, A, A^2, ..., A^{m-1}.

Etant donné que $A^{m+1}=AA^m$, $A^{m+2}=A^2A^m$, ... et que dans la multiplication les puissances des matrices sont commutatives, on aboutit au résultat suivant.

Si le degré du polynôme minimal de A est égal à m, la matrice A^k ($k\geqslant m$) est une combinaison linéaire des matrices E, A, A^2, ..., A^{m-1}:

$$A^k=\sum_{j=0}^{m-1}\varepsilon_{kj}A^j\qquad(k\geqslant m).$$

Considérons maintenant la matrice (16). Puisque la matrice A^kG, où $k\geqslant m$ (et G est une matrice $n\times r$), peut être mise sous la forme

$$A^kG=\sum_{j=0}^{m-1}\varepsilon_{kj}A^jG\qquad(k\geqslant m),$$

le rang de la matrice complétée

$$\widetilde{W}=[G\,AG\,A^2G\,\ldots\,A^{m-1}G\,A^mG\,A^{m+1}G\,\ldots],$$

qui se forme par l'addition à la matrice W des éléments A^mG, $A^{m+1}G$, ..., *reste le même que celui de la matrice W.*

En remplaçant dans (16) A^kG ($k = 0, 1, \ldots, m-1$) par les matrices correspondantes, on obtient la matrice $n \times q$ où

$$q = mr, \tag{11.17}$$

dont les éléments sont des scalaires

$$W = \begin{bmatrix} W_{11} & W_{12} & \ldots & W_{1q} \\ \cdot & \cdot & \cdot & \cdot \\ W_{n1} & W_{n2} & \ldots & W_{nq} \end{bmatrix}. \tag{11.18}$$

Les éléments de la matrice

$$U = \begin{bmatrix} \int_0^T u(\tau)\,\alpha_0(-\tau)\,d\tau \\ \int_0^T u(\tau)\,\alpha_1(-\tau)\,d\tau \\ \cdots \\ \int_0^T u(\tau)\,\alpha_{m-1}(-\tau)\,d\tau \end{bmatrix} \tag{11.19}$$

sont les vecteurs de dimension r, puisque d'après (2)

$$u(\tau) = \begin{bmatrix} u_1(\tau) \\ u_2(\tau) \\ \cdots \\ u_r(\tau) \end{bmatrix}.$$

En introduisant les notations

$$\left.\begin{aligned} U_1 &= \int_0^T u_1(\tau)\,\alpha_0(-\tau)\,d\tau, \\ U_2 &= \int_0^T u_2(\tau)\,\alpha_0(-\tau)\,d\tau, \\ &\cdots\cdots\cdots \\ U_r &= \int_0^T u_r(\tau)\,\alpha_0(-\tau)\,d\tau, \\ U_{r+1} &= \int_0^T u_1(\tau)\,\alpha_1(-\tau)\,d\tau, \\ &\cdots\cdots\cdots \\ U_q &= \int_0^T u_r(\tau)\,\alpha_{m-1}(-\tau)\,d\tau, \end{aligned}\right\} \tag{11.20}$$

on peut mettre la matrice (19) sous la forme

$$U = \begin{bmatrix} U_1 \\ \cdots \\ U_q \end{bmatrix}, \qquad (11.21)$$

c'est-à-dire sous celle d'un vecteur ($q = mr$) de dimension q dont les éléments U_k ($k = 1, \ldots, q$) sont des scalaires.

L'équation vectorielle (15) devient ainsi

$$WU = -x(0), \qquad (11.22)$$

où W et U sont définies par les expressions (18) et (21). L'équation (22) est équivalente au système d'équations scalaires suivant:

$$\left.\begin{array}{l} W_{11}U_1 + W_{12}U_2 + \ldots + W_{1q}U_q = -x_1(0), \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ W_{n1}U_1 + W_{n2}U_2 + \ldots + W_{nq}U_q = -x_n(0). \end{array}\right\} \qquad (11.23)$$

Dans ce qui suit considérons que le nombre de commandes r est choisi de façon à respecter la condition

$$q = mr \geqslant n. \qquad (11.24)$$

Ici, comme nous l'avons indiqué plus haut, m est le degré du polynôme minimal $\psi(\lambda)$ de la matrice A (de plus, $m \leqslant n$), et r, la dimension du vecteur des commandes u.

Si l'on désigne par w_k les vecteurs colonnes de la matrice (18)

$$w_k = \begin{bmatrix} W_{1k} \\ W_{2k} \\ \cdots \\ W_{nk} \end{bmatrix} \qquad (k = 1, \ldots, q), \qquad (11.25)$$

le système d'équations scalaires (23) peut être remplacé par l'équation vectorielle

$$\sum_{k=1}^{q} w_k U_k = -x(0), \qquad (11.26)$$

où w_k ($k = 1, \ldots, q$) et $x(0)$ sont des vecteurs de dimension n et U_k ($k = 1, \ldots, q$) des scalaires; de plus, d'après (24), $q \geqslant n$. Ainsi, le vecteur $x(0)$ est une combinaison linéaire des vecteurs w_k ($k = 1, \ldots, q$).

Comme pour $x(0)$ on peut choisir un vecteur quelconque de l'espace $x_1, \ldots, x_n$ de dimension n, il résulte de la relation (25) que *pour rendre le système* (1) *complètement commandable il faut et il suffit que parmi les vecteurs* w_k ($k = 1, \ldots, q$) *il y ait n vecteurs linéairement indépendants.*

Autrement dit, *la condition de commandabilité* [36] *du système* (1) *est que le rang de la matrice* (16)

$$W = [G\,AG\,A^2G \ldots A^{m-1}G]$$

soit égal à n.

Lorsque $q > n$, le système des variables $U_1, \ldots, U_q$ qui vérifient les équations (23) *n'est pas unique*. Le choix de la solution des équations (23) doit encore être guidé par les relations qu'imposent aux $U_1, \ldots, U_q$ les formules (20). (Ainsi, par exemple, comme il s'ensuit de (20), les fonctions sous le signe de l'intégrale de U_1 et U_{r+1} contiennent le même facteur $u_1(\tau)$ etc.)

Après avoir trouvé $U_1, \ldots, U_q$, il faut, en se servant des relations (20), définir la loi de commande $u_l = u_l(t)$ $(l = 1, \ldots, r)$. La solution de ce problème *n'est pas unique* non plus.

Toutes ces circonstances permettent d'imposer des conditions supplémentaires assurant telle ou telle allure du mouvement du système (conditions d'optimalité, etc).

Montrons maintenant que si l'on connaît les variables $U_1, \ldots, U_q$, il est toujours possible de définir la commande $u(t)$. Cherchons, par exemple, le vecteur fonction $u(t)$, $t \in [0, T]$ sous la forme

$$u(t) = \alpha_0(-t)K_0 + \alpha_1(-t)K_1 + \ldots + \alpha_{m-1}(-t)K_{m-1}, \tag{11.26}$$

où $\alpha_k(-t)$ $(k = 0, 1, \ldots, m-1)$, est un système de fonctions linéairement indépendantes qui font partie de l'expression (12), et $K_0, K_1, \ldots, K_{m-1}$, certains vecteurs de dimension r à rechercher.

Conformément à (19), on a

$$(\alpha_0, \alpha_0)K_0 + (\alpha_1, \alpha_0)K_1 + \ldots + (\alpha_{m-1}, \alpha_0)K_{m-1} = L_0,$$
$$(\alpha_0, \alpha_1)K_0 + (\alpha_1, \alpha_1)K_1 + \ldots + (\alpha_{m-1}, \alpha_1)K_{m-1} = L_1,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(\alpha_0, \alpha_{m-1})K_0 + (\alpha_1, \alpha_{m-1})K_1 + \ldots + (\alpha_{m-1}, \alpha_{m-1})K_{m-1} = L_{m-1}.$$

Ici

$$(\alpha_i, \alpha_j) = \int_0^T \alpha_i(-\tau)\,\alpha_j(-\tau)\,d\tau,$$

et $L_0, L_1, \ldots, L_{m-1}$ désignent les vecteurs de dimension r

$$L_0 = \begin{bmatrix} U_1 \\ \cdots \\ U_r \end{bmatrix}, \quad L_1 = \begin{bmatrix} U_{r+1} \\ \cdots \\ U_{2r} \end{bmatrix}, \quad \ldots, \quad L_{m-1} = \begin{bmatrix} U_{(m-1)r+1} \\ \cdots\cdots \\ U_q \end{bmatrix}$$

où les grandeurs $U_1, \ldots, U_q$ sont fournies par l'équation (25).

Dans ce système d'équations algébriques linéaires le déterminant composé de coefficients affectés aux $K_0, K_1, \ldots, K_{m-1}$ est non

nul du fait que c'est le déterminant de Gram d'un système de fonctions linéairement indépendantes. Ainsi, les vecteurs inconnus $K_0, K_1, \ldots, K_{m-1}$ sont définis d'une façon unique et, par conséquent, il est toujours possible de déterminer la commande $u(t)$ d'après les variables connues $U_1, \ldots, U_q$.

Parmi les commandes possibles du type (26) il y a

$$\tilde{u}(t) = -G^* e^{-A^* t} R^{-1} x(0),$$

où

$$R = \int_0^T e^{-A\tau} GG^* e^{-A^*\tau}\, d\tau,$$

et où l'astérisque désigne la matrice transposée.

Voyons si à l'instant T, $\tilde{u}(t)$ amène le système (1) à l'origine des coordonnées. A cet effet, portons $\tilde{u}(t)$ dans le premier membre de la relation (11)

$$-\int_0^T e^{-A\tau} GG^* e^{-A^*\tau}\, d\tau R^{-1} x(0) = -RR^{-1} x(0) = -x(0).$$

Ainsi la relation (11) est vérifiée, d'où l'on déduit en vertu de (9) que $x(T) = 0$.

Nous avons supposé ici que la matrice R^{-1} existe. Au point 6 nous démontrerons que si le système est commandable, R est une matrice définie positive et, donc, la matrice inverse R^{-1} existe bien.

Exemple. Examinons le système décrit par l'équation différentielle scalaire

$$\frac{dx}{dt} = ax + u.$$

La commande $\tilde{u}(t)$ devient alors

$$\tilde{u}(t) = -\frac{2a}{1 - e^{-2aT}} e^{-at} x(0).$$

Le premier membre de la relation (11) s'écrit maintenant

$$\int_0^T e^{-a\tau} \tilde{u}(\tau)\, d\tau = -\frac{2a}{1 - e^{-2aT}} x(0) \int_0^T e^{-2a\tau}\, d\tau = -x(0),$$

c'est-à-dire la relation (11) est vraie, et d'après (9) on a $x(T) = 0$.

2. Systèmes à une commande. Considérons maintenant un système décrit par les équations différentielles scalaires suivantes:

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} A_{jk} x_k + B_j u \qquad (j = 1, \ldots, n). \qquad (11.27)$$

A la différence du système d'équations (3), le système considéré n'est muni que d'une seule commande u.

En introduisant les matrices

$$x=\begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad A=\begin{bmatrix} A_{11} \ldots A_{1n} \\ \cdots\cdots \\ A_{n1} \ldots A_{nn} \end{bmatrix}, \quad B=\begin{bmatrix} B_1 \\ \cdots \\ B_n \end{bmatrix}, \tag{11.28}$$

on peut remplacer le système (27) par l'équation vectorielle

$$\frac{dx}{dt}=Ax+Bu. \tag{11.29}$$

La matrice $A^{\nu}B$, où ν est un entier quelconque, étant une matrice colonne (vecteur), la condition de commandabilité obtenue au point 1 ne peut être respectée que si le degré du polynôme minimal $\psi(\lambda)$ de la matrice A est égal à n. Ceci a lieu si toutes les valeurs propres de A sont simples. Si parmi ces valeurs il y en a qui sont multiples, le degré m du polynôme minimal $\psi(\lambda)$ ne sera égal à n qu'au cas où à la valeur propre multiple correspond un seul diviseur élémentaire. Le degré de ce diviseur élémentaire est alors égal à l'ordre de multiplicité de la valeur propre.

Si à la valeur propre multiple d'un système correspond non pas un mais plusieurs diviseurs élémentaires, le degré du polynôme minimal est $m < n$. En présence d'une commande unique, la condition de commandabilité de ces systèmes n'est pas observée.

Pour les systèmes dans lesquels

$$m = n, \tag{11.30}$$

la matrice (16) se met sous la forme

$$Q = [B\,AB\,A^2B \ldots A^{n-1}B]. \tag{11.31}$$

Les éléments de la matrice (31) sont des vecteurs de dimension n et la condition de commandabilité du système est que le rang de la matrice Q soit égal à n.

Le vecteur (21) s'écrit alors

$$V=\begin{bmatrix} \int_0^T u(\tau)\,\alpha_0(-\tau)\,d\tau \\ \int_0^T u(\tau)\,\alpha_1(-\tau)\,d\tau \\ \cdots\cdots\cdots\cdots \\ \int_0^T u(\tau)\,\alpha_{n-1}(-\tau)\,d\tau \end{bmatrix}. \tag{11.32}$$

Les éléments du vecteur V sont des scalaires.

Dans le cas considéré, l'équation (22) devient

$$QV = -x(0). \qquad (11.33)$$

Si la condition de commandabilité est vérifiée, c'est-à-dire si le rang de la matrice Q est égal à n, le déterminant de Q est différent du zéro et l'équation (33) donne

$$V = -Q^{-1}x(0), \qquad (11.34)$$

où Q^{-1} désigne la matrice inverse.

Conformément à (32), la relation vectorielle (34) est équivalente à n relations scalaires

$$\left.\begin{array}{l} \int_0^T u(\tau)\,\alpha_0(-\tau)\,d\tau = \varkappa_1, \\ \int_0^T u(\tau)\,\alpha_1(-\tau)\,d\tau = \varkappa_2, \\ \ldots\ldots\ldots\ldots\ldots \\ \int_0^T u(\tau)\,\alpha_{n-1}(-\tau)\,d\tau = \varkappa_n, \end{array}\right\} \qquad (11.35)$$

où $\varkappa_i = \text{const}$ $(i = 1, \ldots, n)$ sont les éléments du vecteur $-Q^{-1}x(0)$.

La commande $u = u(t)$ doit être choisie de façon à observer les conditions (35) qui ne déterminent pourtant pas d'une façon univoque la fonction $u(t)$.

3. Observabilité des systèmes décrits par les équations différentielles linéaires à coefficients constants. Examinons le système décrit par les équations différentielles vectorielles

$$\frac{d\tilde{x}}{dt} = A\tilde{x} + Gu, \qquad (11.36)$$

$$\tilde{y} = C\tilde{x} + Lu. \qquad (11.37)$$

Ici $\tilde{x}$ est le vecteur de dimension n ; $\tilde{y}$, le vecteur de dimension p ; u, le vecteur de dimension r. Par A, G, C et L on désigne les matrices $n \times n$; $n \times r$; $p \times n$; $p \times r$ respectivement.

Le vecteur $\tilde{x}$ est le vecteur d'état du système (ses composantes $\tilde{x}_1, \ldots, \tilde{x}_n$ sont les coordonnées de phase du système) ; u, le vecteur dont les composantes sont des commandes.

Désignons par $\tilde{y}$ le vecteur

$$\tilde{y} = \begin{bmatrix} \tilde{y}_1 \\ \cdots \\ \tilde{y}_p \end{bmatrix},$$

dont les éléments $\tilde{y}_1, \ldots, \tilde{y}_p$ sont des combinaisons linéaires des coordonnées de phase $\tilde{x}_k$ $(k = 1, \ldots, n)$ et des commandes u_l $(l = 1, \ldots, r)$.

Supposons ensuite que dans l'intervalle de temps $0 \leqslant t \leqslant T$ les éléments $\tilde{y}_1, \ldots, \tilde{y}_p$ du vecteur $\tilde{y}$ sont accessibles à l'observation et ainsi, d'après la mesure, on connaît les fonctions $\tilde{y}_1(t), \ldots, \tilde{y}_p(t)$ dans l'intervalle de temps $0 \leqslant t \leqslant T$. Supposons également connue la loi de variation des commandes $u_l = u_l(t)$ dans l'intervalle de temps considéré $0 \leqslant t \leqslant T$.

Puisque d'après (7)

$$\tilde{x}(t) = e^{At}\tilde{x}(0) + \int_0^t e^{A(t-\tau)}Gu(\tau)\,d\tau,$$

on a en vertu de (37)

$$\tilde{y}(t) = Ce^{At}\tilde{x}(0) + C\int_0^t e^{A(t-\tau)}Gu(\tau)\,d\tau + Lu(t). \tag{11.38}$$

Comme la fonction $u(t)$ est supposée connue, le deuxième et le troisième terme du second membre de (38) peuvent être calculés et ces termes peuvent être retranchés de la fonction $\tilde{y}(t)$ obtenue par observation:

$$\tilde{y}(t) - C\int_0^t e^{A(t-\tau)}Gu(\tau)\,d\tau - Lu(t) = Ce^{At}\tilde{x}(0). \tag{11.39}$$

Nous connaissons donc le premier membre de la relation (39). L'état initial du système $\tilde{x}(0)$ est supposé inconnu. On se demande si l'on peut obtenir la valeur $\tilde{x}(0)$ d'après les données fournies par les observations exposées ci-dessus.

Le problème que nous venons de formuler est équivalent au problème suivant. Ramener le système décrit par les équations vectorielles (qui diffèrent des équations (36) et (37) du fait que $u(t) \equiv 0$):

$$\frac{dx}{dt} = Ax, \tag{11.40}$$

$$y = Cx \tag{11.41}$$

à la valeur initiale $x(0)$ en partant du vecteur fonction $y(t)$ $(0 \leqslant t \leqslant T)$ établi par observation.

La possibilité de ramener le système à l'état initial $x(0)$ en partant de la sortie observée s'appelle *observabilité*.

Les systèmes qui jouissent de cette propriété sont dits *complètement observables*.

Les équations (40) et (41) entraînent

$$y(t) = Ce^{At}x(0). \tag{11.42}$$

La matrice C qui fait partie de l'équation (41) est de la forme:

$$C = \begin{bmatrix} c_{11} & c_{12} \dots c_{1n} \\ c_{21} & c_{22} \dots c_{2n} \\ \dots & \dots \\ c_{p1} & c_{p2} \dots c_{pn} \end{bmatrix}. \tag{11.43}$$

En introduisant pour les lignes de la matrice (43) la notation C_j ($j = 1, \dots, p$)

$$C_j = [c_{j1} c_{j2} \dots c_{jn}] \quad (j = 1, \dots, p), \tag{11.44}$$

on peut la mettre sous la forme partitionnée

$$C = \begin{bmatrix} C_1 \\ \dots \\ C_p \end{bmatrix}. \tag{11.45}$$

En vertu de (42), l'élément $y_j(t)$ du vecteur $y(t)$ peut s'écrire

$$y_j(t) = C_j e^{At}x(0) \quad (j = 1, \dots, p). \tag{11.46}$$

D'après (10.58)

$$e^{At} = \sum_{k=0}^{m-1} \alpha_k(t) A^k, \tag{11.47}$$

où m est le degré du polynôme minimal $\psi(\lambda)$ de la matrice A, et $\alpha_k(t)$ sont les coefficients du polynôme d'interpolation de Lagrange-Sylvester $r(\lambda)$ construit pour la fonction $e^{\lambda t}$, définie dans le spectre de A. C'est pourquoi l'expression (46) peut être ramenée à la forme

$$y_j(t) = \sum_{k=0}^{m-1} \alpha_k(t) C_j A^k x(0) \quad (j = 1, \dots, p). \tag{11.48}$$

En multipliant les premier et second membres de la relation (48) par $\alpha_i(t)$ ($i = 0, 1, \dots, m-1$) et en intégrant les fonctions obtenues par rapport à t dans les limites de 0 à T, on obtient

$$\int_0^T \alpha_i(t) y_j(t)\, dt = \sum_{k=0}^{m-1} \int_0^T \alpha_i(t) \alpha_k(t)\, dt\, C_j A^k x(0) \tag{11.49}$$

$$(j = 1, \dots, p;\ i = 0, 1, \dots, m-1).$$

Introduisons maintenant les notations:

$$(\varphi_i, \psi_j) = \int_0^T \varphi_i(t) \psi_j(t)\, dt. \tag{11.50}$$

En vertu de (49), on a pour tout j fixé le système d'équations suivant :

$$\left.\begin{aligned}
&(\alpha_0, \alpha_0)\, C_j x(0) + (\alpha_0, \alpha_1)\, C_j A x(0) + \ldots + \\
&\qquad + (\alpha_0, \alpha_{m-1})\, C_j A^{m-1} x(0) = (\alpha_0, y_j), \\
&(\alpha_1, \alpha_0)\, C_j x(0) + (\alpha_1, \alpha_1)\, C_j A x(0) + \ldots + \\
&\qquad + (\alpha_1, \alpha_{m-1})\, C_j A^{m-1} x(0) = (\alpha_1, y_j), \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&(\alpha_{m-1}, \alpha_0)\, C_j x(0) + (\alpha_{m-1}, \alpha_1)\, C_j A x(0) + \ldots + \\
&\qquad + (\alpha_{m-1}, \alpha_{m-1})\, C_j A^{m-1} x(0) = (\alpha_{m-1}, y_j).
\end{aligned}\right\} \quad (11.51)$$

Le déterminant composé de coefficients affectés aux $C_j x(0)$, $C_j A x(0)$, ..., $C_j A^{m-1} x(0)$ des équations (51) est différent du zéro

$$\Delta = \begin{vmatrix} (\alpha_0, \alpha_0) & (\alpha_0, \alpha_1) & \ldots & (\alpha_0, \alpha_{m-1}) \\ (\alpha_1, \alpha_0) & (\alpha_1, \alpha_1) & \ldots & (\alpha_1, \alpha_{m-1}) \\ \ldots & \ldots & \ldots & \ldots \\ (\alpha_{m-1}, \alpha_0) & (\alpha_{m-1}, \alpha_1) & \ldots & (\alpha_{m-1}, \alpha_{m-1}) \end{vmatrix} \neq 0, \quad (11.52)$$

du fait que c'est le déterminant de Gram [23, p. 79] d'un système de fonctions linéairement indépendantes $\alpha_0(t)$, $\alpha_1(t)$, ..., $\alpha_{m-1}(t)$.

Ainsi le système d'équations (52) possède une seule solution

$$\left.\begin{aligned}
C_j x(0) &= \mu_{j1}, \\
C_j A x(0) &= \mu_{j2}, \\
&\ldots\ldots \\
C_j A^{m-1} x(0) &= \mu_{jm}.
\end{aligned}\right\} \quad (11.53)$$

Etant donné que $j = 1, \ldots, p$, on a ν relations du type (53), où

$$\nu = mp. \quad (11.54)$$

Introduisons maintenant les notations suivantes pour le produit scalaire de deux vecteurs ξ et η :

$$\langle \xi, \eta \rangle = \xi_1\eta_1 + \xi_2\eta_2 + \ldots + \xi_n\eta_n. \quad (11.55)$$

Désignons par M^* la transposée de la matrice M. Puisque la transposition du produit de deux matrices M et N donne lieu à la relation

$$(MN)^* = N^*M^*, \quad (11.56)$$

les équations (53) peuvent s'écrire

$$\left.\begin{aligned}
&\langle C_j^*, x(0)\rangle = \mu_{j1}, \\
&\langle A^*C_j^*, x(0)\rangle = \mu_{j2}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
&\langle A^{*m-1}C_j^*, x(0)\rangle = \mu_{jm} \qquad (j = 1, \ldots, p).
\end{aligned}\right\} \quad (11.57)$$

Conformément à (44) on a

$$C_j^* = \begin{bmatrix} c_{j1} \\ c_{j2} \\ \cdots \\ c_{jn} \end{bmatrix}. \qquad (11.58)$$

Les matrices C^*, A^*C^*, $A^{*2}C^*$, ..., $A^{*m-1}C^*$ seront

$$C^* = \begin{bmatrix} c_{11} & c_{21} & \dots & c_{p1} \\ c_{12} & c_{22} & \dots & c_{p2} \\ \cdot & \cdot & \cdot & \cdot \\ c_{1n} & c_{2n} & \dots & c_{pn} \end{bmatrix} = [C_1^* \; C_2^* \dots C_p^*],$$

$$A^*C^* = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix} \times \begin{bmatrix} c_{11} & c_{21} & \dots & c_{p1} \\ c_{12} & c_{22} & \dots & c_{p2} \\ \cdot & \cdot & \cdot & \cdot \\ c_{1n} & c_{2n} & \dots & c_{pn} \end{bmatrix} =$$

$$= [A^*C_1^* \; A^*C_2^* \dots A^*C_p^*],$$

$$A^{*2}C^* = [A^{*2}C_1^* \; A^{*2}C_2^* \dots A^{*2}C_p^*],$$

$$\dots\dots\dots\dots\dots\dots \qquad (11.59)$$

$$A^{*m-1}C^* = [A^{*m-1}C_1^* \; A^{*m-1}C_2^* \dots A^{*m-1}C_p^*].$$

On voit de (59) et (57) que les colonnes des matrices C^*, A^*C^*, $A^{*2}C^*$, ..., $A^{*m-1}C^*$ (le nombre de ces colonnes est $\nu = mp$) sont des vecteurs qui interviennent dans les premiers membres des relations (57). Les vecteurs concernés sont ainsi des colonnes de la matrice partitionnée suivante

$$S = [C^* A^*C^* A^{*2}C^* \dots A^{*m-1}C^*]. \qquad (11.60)$$

Remarquons que si les blocs de la matrice (60) sont remplacés par les matrices correspondantes, on obtient une matrice $n \times \nu$ (où $\nu = mp$) dont les éléments sont des scalaires.

Supposons maintenant que parmi les colonnes de la matrice S il y a n colonnes S_{σ_1}, S_{σ_2}, ..., S_{σ_n} linéairement indépendantes, c'est-à-dire le rang de S est n.

Supposons encore que la dimension p du vecteur y est choisie de façon à respecter la condition

$$\nu = mp \geqslant n. \qquad (11.61)$$

Choisissons des ν équations (57) celles dont font partie les vecteurs $S_{\sigma_1}, S_{\sigma_2}, \ldots, S_{\sigma_n}$. Ces équations sont

$$\left.\begin{array}{l}\langle S_{\sigma_1}, x(0)\rangle = \mu_{\sigma_1},\\ \langle S_{\sigma_2}, x(0)\rangle = \mu_{\sigma_2},\\ \dots\dots\dots\dots\\ \langle S_{\sigma_n}, x(0)\rangle = \mu_{\sigma_n},\end{array}\right\} \qquad (11.62)$$

ou

$$\left.\begin{array}{l}S_{\sigma_1 1}x_1(0) + S_{\sigma_1 2}x_2(0) + \ldots + S_{\sigma_1 n}x_n(0) = \mu_{\sigma_1},\\ S_{\sigma_2 1}x_1(0) + S_{\sigma_2 2}x_2(0) + \ldots + S_{\sigma_2 n}x_n(0) = \mu_{\sigma_2},\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ S_{\sigma_n 1}x_1(0) + S_{\sigma_n 2}x_2(0) + \ldots + S_{\sigma_n n}x_n(0) = \mu_{\sigma_n}.\end{array}\right\} \qquad (11.63)$$

Nous avons obtenu un système d'équations algébriques linéaires inhomogènes par rapport à $x_1(0), x_2(0), \ldots, x_n(0)$. Le déterminant composé de coefficients des équations (63) est non nul puisque les vecteurs $S_{\sigma_1}, S_{\sigma_2}, \ldots, S_{\sigma_n}$ sont linéairement indépendants et, par suite, le système d'équations (63) a une solution unique, c'est-à-dire qu'il détermine d'une façon unique les composantes $x_1(0), x_2(0), \ldots$ $\ldots, x_n(0)$ du vecteur de l'état initial du système $x(0)$.

On peut encore donner aux relations (62) l'interprétation suivante. Les vecteurs linéairement indépendants $S_{\sigma_1}, S_{\sigma_2}, \ldots, S_{\sigma_n}$ peuvent être considérés comme base de l'espace $x_1, \ldots, x_n$ de dimension n. Les grandeurs $\mu_{\sigma_1}, \mu_{\sigma_2}, \ldots, \mu_{\sigma_n}$ sont des projections du vecteur $x(0)$ sur les vecteurs de base. Ces projections déterminent d'une façon unique le vecteur $x(0)$ recherché.

Ainsi *la condition nécessaire et suffisante d'observabilité du système* (36), (37) *consiste à égaler à n le rang de la matrice* (60)

$$S = [C^* \; A^*C^* \; A^{*2}C^* \; \ldots \; A^{*\,m-1}C^*].$$

R e m a r q u e s u r l'i n o b s e r v a b i l i t é. Supposons maintenant que le rang r de la matrice S soit plus petit que n. Désignons par $S^{[k]}$ la k-ième colonne de la matrice S et écrivons cette matrice sous la forme

$$S = \begin{bmatrix} S_1^{[1]} & S_1^{[2]} \ldots S_1^{[\nu]} \\ S_2^{[1]} & S_2^{[2]} \ldots S_2^{[\nu]} \\ \dots & \dots\dots\dots \\ S_n^{[1]} & S_n^{[2]} \ldots S_n^{[\nu]} \end{bmatrix}.$$

Notons que d'après (61) le nombre de colonnes de S

$$\nu \geqslant n.$$

On voit facilement que dans le cas de $r < n$, il existe des vecteurs $x(0) \neq 0$ qui vérifient les conditions

$$\langle S^{[k]}, x(0)\rangle = 0 \qquad (k = 1, \ldots, \nu). \qquad (*)$$

En effet, les relations (*) peuvent s'écrire

$$\begin{aligned} S_1^{[1]}x_1(0)+S_2^{[1]}x_2(0)+\ldots+S_n^{[1]}x_n(0)&=0,\\ S_1^{[2]}x_1(0)+S_2^{[2]}x_2(0)+\ldots+S_n^{[2]}x_n(0)&=0,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots&\\ S_1^{[\nu]}x_1(0)+S_2^{[\nu]}x_2(0)+\ldots+S_n^{[\nu]}x_n(0)&=0. \end{aligned} \qquad (**)$$

Nous avons là un système composé de ν (où $\nu \geqslant n$) équations algébriques linéaires homogènes par rapport à n inconnues $x_1(0)$, $x_2(0)$, . . ., $x_n(0)$. La matrice des coefficients affectés à $x_1(0)$, $x_2(0)$, . . ., $x_n(0)$ dans les équations (**) est la matrice S^*, et comme le rang r de S est inférieur à n, le système (**) a les solutions autres que la solution triviale

$$\tilde{x}(0)=\begin{bmatrix}\tilde{x}_1(0)\\ \tilde{x}_2(0)\\ \cdots\\ \tilde{x}_n(0)\end{bmatrix} \neq 0.$$

Les colonnes de la matrice S étant les colonnes des matrices C^*, A^*C^*, ..., $A^{*m-1}C^*$, les relations (*) entraînent que pour les vecteurs $x(0)$ fournis par les équations (**)

$$C_jA^k\tilde{x}(0)=0 \qquad (j=1,\ \ldots,\ p;\ k=0,\ 1,\ \ldots,\ m-1). \qquad (***)$$

De plus, d'après (48),

$$y_j(t)=\sum_{k=0}^{m-1}\alpha_k(t)\,C_jA^k\tilde{x}(0) \qquad (j=1,\ \ldots,\ p),$$

c'est-à-dire

$$y_j(t)\equiv 0 \qquad 0\leqslant t<\infty \qquad (j=1,\ \ldots,\ p).$$

Ainsi dans le cas où le rang r de S est inférieur à n, il existe des états initiaux du système $\tilde{x}(0)\neq 0$ non uniques et non nuls tels que le vecteur fonction $y(t)\equiv 0$, $0\leqslant t<\infty$. Ceci traduit précisément l'*inobservabilité* d'un système.

Remarquons que $y(t)\equiv 0$, $0\leqslant t<\infty$ également dans le cas où l'état initial du système est nul $x(0)=0$.

4. Systèmes à une coordonnée observable. Considérons maintenant un système décrit par les équations

$$\frac{dx_j}{dt}=\sum_{k=1}^{n}A_{jk}x_k \quad (j=1,\ \ldots,\ n). \qquad (11.64)$$

$$y=\sum_{k=1}^{n}c_kx_k. \qquad (11.65)$$

A la différence des équations (40), (41), les observations ne renseignent ici que sur la fonction scalaire $y=y(t)$ déterminée par l'ex-

pression (65), et non sur le vecteur fonction, comme c'était le cas du point 3.

En introduisant les matrices

$$x=\begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad A=\begin{bmatrix} A_{11} \dots A_{1n} \\ \cdots\cdots\cdots \\ A_{n1} \dots A_{nn} \end{bmatrix}, \quad C=[c_1, c_2 \dots c_n], \quad (11.66)$$

on peut récrire les équations (64) et (65) sous la forme:

$$\frac{dx}{dt}=Ax, \quad (11.67)$$

$$y=Cx. \quad (11.68)$$

Conformément à (66) la matrice C^* étant une matrice colonne (vecteur)

$$C^*=\begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix}, \quad (11.69)$$

la matrice $A^{*k}C^*$, où k est un entier quelconque, est une matrice colonne (vecteur). C'est pourquoi la condition d'observabilité du point 3 ne peut être respectée ici que si le degré du polynôme minimal $\psi(\lambda)$ de la matrice A est égal à n. Ceci n'est possible que si à chaque valeur propre multiple de A correspond un seul diviseur élémentaire.

Pour les systèmes à une coordonnée observable, dans lesquels

$$m=n, \quad (11.70)$$

la matrice (60) s'écrit

$$P=[C^* \, A^*C^* \, A^{*2}C^* \, \dots \, A^{*n-1}C^*]. \quad (11.71)$$

Les éléments de la matrice (71) sont des vecteurs de dimension n et la condition d'observabilité du système est que le rang de la matrice P soit égal à n.

Le problème considéré donne lieu à n équations suivantes du type (57)

$$\left.\begin{aligned} \langle C^*, x(0)\rangle &= \mu_1, \\ \langle A^*C^*, x(0)\rangle &= \mu_2, \\ \langle A^{*2}C^*, x(0)\rangle &= \mu_3, \\ &\cdots\cdots\cdots \\ \langle A^{*n-1}C^*, x(0)\rangle &= \mu_n. \end{aligned}\right\} \quad (11.72)$$

Si les vecteurs C^*, A^*C^*, ..., $A^{*n-1}C^*$ sont linéairement indépendants, ils forment la base de l'espace $x_1, \dots, x_n$ de dimension n.

Les grandeurs $\mu_1, \mu_2, \ldots, \mu_n$ constituent alors les projections du vecteur $x(0)$ sur les vecteurs de base. Ces projections définissent le vecteur $x(0)$ de façon unique.

5. Dualité dans la théorie de commandabilité et d'observabilité. Examinons le système I décrit par les équations

$$\left.\begin{aligned} \frac{dx}{dt} &= Ax + Gu, \\ y &= Cx + Lu. \end{aligned}\right\} \tag{11.73}$$

Ici x, u, y sont les vecteurs de dimension n, r et p, et A, G, C, L, les matrices $n \times n$, $n \times r$, $p \times n$, $p \times r$.

La condition de commandabilité est que le rang de la matrice

$$W = [G\,AG\,A^2G \ldots A^{m-1}G]$$

soit égal à n.

La condition d'observabilité est que le rang de la matrice

$$S = [C^*\,A^*C^*\,A^{*2}C^* \ldots A^{*m-1}C^*]$$

soit égal à n.

Soit le système II décrit par les équations

$$\left.\begin{aligned} \frac{d\xi}{dt} &= A^*\xi + C^*v, \\ \eta &= G^*\xi + L^*v. \end{aligned}\right\} \tag{11.74}$$

Ici ξ, v, η sont des vecteurs de dimension n, p et r.

Pour le système II la condition d'observabilité est que le rang de la matrice

$$S = [C^*\,A^*C^*\,A^{*2}C^* \ldots A^{*m-1}C^*]$$

soit égal à n.

La condition d'observabilité du système II est que le rang de la matrice

$$W = [G\,AG\,A^2G \ldots A^{m-1}G]$$

soit égal à n.

Nous sommes donc ici dans le cas du principe de dualité dans le problème de commandabilité et d'observabilité, établi par Kalman [36].

Le système I *est commandable* (observable) *si et seulement si le système* II *est observable* (respectivement commandable).

6. Commandabilité des systèmes linéaires non stationnaires. Examinons le système linéaire non stationnaire décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = A(t)\,x + G(t)\,u, \tag{11.75}$$

où

$$\left.\begin{aligned} x&=\begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad A(t)=\begin{bmatrix} A_{11}(t) \ldots A_{1n}(t) \\ \cdots\cdots\cdots\cdots \\ A_{n1}(t) \ldots A_{nn}(t) \end{bmatrix}, \\ u&=\begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad G(t)=\begin{bmatrix} G_{11}(t) \ldots G_{1r}(t) \\ \cdots\cdots\cdots\cdots \\ G_{n1}(t) \ldots G_{nr}(t) \end{bmatrix}. \end{aligned}\right\} \tag{11.76}$$

Les éléments des matrices $A(t)$ et $G(t)$ sont des fonctions réelles continues du temps.

Le système (75) est dit *complètement commandable à l'instant* t_0 si l'on peut le faire passer *en un intervalle de temps fini* de l'état où il se trouve à l'instant t_0 à l'état nul en choisissant dûment la loi de variation des commandes $u_l = u_l(t)$ $(l = 1, \ldots, r)$.

Les propriétés d'un système non stationnaire se modifiant dans le temps, il existe des systèmes commandables (au sens de la définition donnée ci-dessus) à l'instant T et non commandables à n'importe quel autre instant ultérieur. La définition même de la commandabilité implique que si le système est commandable à l'instant T, il l'est également à n'importe quel instant $t_0 < T$ (du fait qu'il sera commandable à l'instant T, quel que soit son état).

Théorème. *Soit* $\Phi(t, \tau) = \theta(t)\,\theta^{-1}(\tau)$, *où* $\theta(t)$ *est la matrice fondamentale des solutions du système décrit par l'équation différentielle vectorielle homogène*

$$\frac{dx}{dt} = A(t)\,x, \tag{11.77}$$

et $W(t_0, t)$ *la matrice*

$$W(t_0, t) = \int_{t_0}^{t} \Phi(t_0, \tau)\,G(\tau)\,G^*(\tau)\,\Phi^*(t_0, \tau)\,d\tau. \tag{11.78}$$

Le système linéaire non stationnaire (75) *est commandable à l'instant* t_0 *si et seulement si pour un certain* t_1 *fini la matrice* $W(t_0, t_1)$ *déterminée par l'expression* (78) *est une matrice définie positive* *.

Démonstration.

1°. Par hypothèse, la matrice $W(t_0, t_1)$ est une matrice symétrique définie positive. Par conséquent, c'est une matrice régulière, c'est-à-dire $\det W(t_0, t_1) \neq 0$.

Ainsi l'inverse $W^{-1}(t_0, t_1)$ existe bien et le vecteur des commandes $u(t)$ peut être choisi sous la forme:

$$\tilde{u}(t) = -G^*(t)\,\Phi^*(t_0, t)\,W^{-1}(t_0, t_1)\,x(t_0). \tag{11.79}$$

---

* La matrice R $(n \times n)$ est dite définie positive si la forme quadratique $\langle \xi, R\xi \rangle$ (où ξ est un vecteur de dimension n) est positive pour tout $\xi \neq 0$.

Puisque d'après (7.26)

$$x(t) = \Phi(t, t_0)\,x(t_0) + \int_{t_0}^{t} \Phi(t, \tau)\,G(\tau)\,\tilde{u}(\tau)\,d\tau, \qquad (11.80)$$

à l'instant $t = t_1$ l'état du système est

$$x(t_1) = \Phi(t_1, t_0)\,x(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \tau)\,G(\tau)\,\tilde{u}(\tau)\,d\tau. \qquad (11.81)$$

En portant dans (81) l'expression (79) pour $\tilde{u}(t)$, on obtient

$$x(t_1) = \Phi(t_1, t_0)\,x(t_0) -$$
$$- \int_{t_0}^{t_1} \Phi(t_1, \tau)\,G(\tau)\,G^*(\tau)\,\Phi^*(t_0, \tau)\,d\tau\,W^{-1}(t_0, t_1)\,x(t_0). \qquad (11.82)$$

Nous avons montré plus haut (7.24) et (7.25) que

$$\Phi(t_1, t_0)\,\Phi(t_0, t_1) = E,$$
$$\Phi(t_0, t_1)\,\Phi(t_1, \tau) = \Phi(t_0, \tau).$$

Donc, on a, compte tenu de (78),

$$\int_{t_0}^{t_1} \Phi(t_1, \tau)\,G(\tau)\,G^*(\tau)\,\Phi^*(t_0, \tau)\,d\tau =$$
$$= \Phi(t_1, t_0) \int_{t_0}^{t_1} \Phi(t_0, t_1)\,\Phi(t_1, \tau)\,G(\tau)\,G^*(\tau)\,\Phi^*(t_0, \tau)\,d\tau =$$
$$= \Phi(t_1, t_0)\,W(t_0, t_1). \qquad (11.83)$$

En vertu de (82) et (83), l'état du système à l'instant $t = t_1$ s'écrit

$$x(t_1) = \Phi(t_1, t_0)\,x(t_0) - \Phi(t_1, t_0)\,W(t_0, t_1)\,W^{-1}(t_0, t_1)\,x(t_0) = 0. \qquad (11.84)$$

Ainsi, à l'instant $t = t_1$, le système (75) est en effet recalé à zéro.

Remarquons que la commande $\tilde{u}(t)$ donnée par l'expression (79) qui, comme nous l'avons démontré, ramène le système (75) à zéro à l'instant t_1, *n'est pas unique*. En effet, en vertu de (81), à l'instant $t = t_1$ la commande

$$u(t) = \tilde{u}(t) + v(t),$$

où $v(t)$ est un vecteur fonction quelconque de dimension r vérifiant la condition

$$\int_{t_0}^{t_1} \Phi(t_1, \tau)\,G(\tau)\,v(\tau)\,d\tau = 0,$$

amène également le système (75) à zéro.

D'une façon analogue à l'instant $t = t_1$ la commande

$\hat{u}(t) = -G^*(t)\,\Phi^*(t_0, t)\,W^{-1}(t_0, t_1)\,[x(t_0) - \Phi(t_0, t_1)\,x^*]$

amène le système au point $x(t_1) = x^*$.

2°. Montrons maintenant que si le système (75) est commandable, la matrice $W(t_0, t_1)$ est définie positive.

Considérons d'abord la forme quadratique $\langle \xi, W(t_0, t_1)\,\xi \rangle$, où ξ est un vecteur de dimension n. D'après (78), on a

$$\langle \xi, W\xi \rangle = \int_{t_0}^{t_1} \langle \xi, \Phi G G^* \Phi^* \xi \rangle \, dt. \qquad (11.85)$$

Etant donné que pour toute matrice rectangulaire a du type $n \times r$

$$\langle \xi, aa^*\xi \rangle = \xi^* aa^* \xi,$$

$$(a^*\xi)^* = \xi^* a,$$

$$\langle a^*\xi, a^*\xi \rangle = (a^*\xi)^*\, a^*\xi = \xi^* aa^* \xi,$$

il vient

$$\langle \xi, aa^*\xi \rangle = \langle a^*\xi, a^*\xi \rangle. \qquad (11.86)$$

Ainsi, conformément à (86) et (85), tout ξ donne lieu à la relation

$$\langle \xi, W\xi \rangle = \int_{t_0}^{t_1} \langle G^*\Phi^*\xi, G^*\Phi^*\xi \rangle \, dt \geqslant 0. \qquad (11.87)$$

Pour achever la démonstration, il faut encore prouver que la matrice W est régulière; la relation (87) acquiert alors la forme requise $\langle \xi, W\xi \rangle > 0$.

Raisonnons par l'absurde: soit W une matrice singulière. Il existe alors un vecteur $\nu \neq 0$ tel que

$$\langle \nu, W\nu \rangle = 0. \qquad (11.88)$$

Désignons maintenant par $\varkappa(t)$ le vecteur

$$\varkappa(t) = -G^*(t)\,\Phi^*(t_0, t)\,\nu. \qquad (11.89)$$

En vertu de (89), $\varkappa(t)$ est une fonction continue de t. Conformément à (89), (86), (78) et (88), on a

$$\int_{t_0}^{t_1} \langle \varkappa(t), \varkappa(t) \rangle \, dt = \int_{t_0}^{t_1} \langle G^*\Phi^*\nu, G^*\Phi^*\nu \rangle \, dt =$$

$$= \int_{t_0}^{t_1} \langle \nu, \Phi G G^* \Phi^* \nu \rangle \, dt = \left\langle \nu, \int_{t_0}^{t_1} \Phi G G^* \Phi^* \, dt\, \nu \right\rangle = \langle \nu, W\nu \rangle = 0 \qquad (11.90)$$

d'où il résulte

$$\varkappa(t) \equiv 0 \qquad (t_0 \leqslant t \leqslant t_1). \qquad (11.91)$$

Retenons maintenant que par hypothèse le système concerné est commandable. Par conséquent, il existe une certaine commande $u(t)$ qui amène le système de l'état $\nu \neq 0$ (à l'instant t_0) à l'état nul (à l'instant t_1). D'une façon analogue à (81), on a la relation

$$0 = \Phi(t_1, t_0)\nu + \int_{t_0}^{t_1} \Phi(t_1, t) G(t) u(t) dt. \qquad (11.92)$$

Il en résulte que

$$\nu = -\Phi(t_0, t_1) \int_{t_0}^{t_1} \Phi(t_1, t) G(t) u(t) dt,$$

ou

$$\nu = -\int_{t_0}^{t_1} \Phi(t_0, t) G(t) u(t) dt. \qquad (11.93)$$

On tire de (91) et de (89) que

$$0 = \int_{t_0}^{t_1} \langle \varkappa(t), u(t) \rangle dt = -\int_{t_0}^{t_1} \langle G^*(t) \Phi^*(t_0, t)\nu, u(t) \rangle dt. \qquad (11.94)$$

Puisque pour toute matrice rectangulaire a on a

$$\langle a^*\nu, u \rangle = (a^*\nu)^* u = \nu^* a u = \langle \nu, au \rangle,$$

la relation (94) entraîne

$$-\int_{t_0}^{t_1} \langle \nu, \Phi(t_0, t) G(t) u(t) \rangle dt =$$

$$= \left\langle \nu, -\int_{t_0}^{t_1} \Phi(t_0, t) G(t) u(t) dt \right\rangle = 0. \qquad (11.95)$$

D'après (93) la relation (95) se met sous la forme

$$\langle \nu, \nu \rangle = 0, \qquad (11.96)$$

ce qui contredit l'hypothèse initiale suivant laquelle $\nu \neq 0$.

Cette contradiction résulte de l'hypothèse (88) que la matrice $W(t_0, t_1)$ est singulière. Nous avons donc établi que $W(t_0, t_1)$ est une matrice régulière, c'est-à-dire la relation (88) n'est vraie pour aucun vecteur $\nu \neq 0$ et, par conséquent,

$$\langle \nu, W\nu \rangle \neq 0. \qquad (11.97)$$

Les relations (87) et (97) entraînent que

$$\langle \xi, W\xi \rangle > 0, \qquad (11.98)$$

c'est-à-dire si le système (75) est commandable, $W(t_0, t_1)$ est une matrice définie positive.

7. Observabilité des systèmes linéaires non stationnaires. Considérons maintenant le problème qui consiste à restituer la valeur initiale $x(t_0)$ d'après le vecteur fonction $y(t)$ $(t_0 \leqslant t \leqslant T)$ fourni par les observations pour le système

$$\frac{dx}{dt} = A(t)\,x, \tag{11.99}$$

$$y(t) = C(t)\,x. \tag{11.100}$$

Ici x est le vecteur de dimension n; y, le vecteur de dimension p; $A(t)$, la matrice $n \times n$; $C(t)$, la matrice $p \times n$. Les éléments des matrices $A(t)$ et $C(t)$ sont des fonctions continues réelles du temps.

Si tout état initial $x(t_0)$ peut être déterminé d'après le vecteur fonction $y(t)$ connu dans l'intervalle $[t_0, T]$, dans cet intervalle le système est dit complètement observable.

D'après (80) et (100)

$$x(t) = \Phi(t, t_0)\,x(t_0), \tag{11.101}$$

$$y(t) = C(t)\,\Phi(t, t_0)\,x(t_0). \tag{11.102}$$

En multipliant à gauche le premier et le deuxième membre de (102) par la matrice $\Phi^*(t, t_0)\,C^*(t)$ et en intégrant par rapport à t de t_0 à T, on obtient

$$\int_{t_0}^{T} \Phi^*(t, t_0)\,C^*(t)\,y(t)\,dt = \int_{t_0}^{T} \Phi^*(t, t_0)\,C^*(t)\,C(t)\,\Phi(t, t_0)\,dt\,x(t_0). \tag{11.103}$$

Introduisons les notations

$$M(t_0, T) = \int_{t_0}^{T} \Phi^*(t, t_0)\,C^*(t)\,C(t)\,\Phi(t, t_0)\,dt, \tag{11.104}$$

$$\eta(t_0, T) = \int_{t_0}^{T} \Phi^*(t, t_0)\,C^*(t)\,y(t)\,dt, \tag{11.105}$$

pour écrire la relation (103) sous la forme

$$M(t_0, T)\,x(t_0) = \eta(t_0, T). \tag{11.106}$$

Si la matrice $M(t_0, T)$ est régulière, c'est-à-dire si son déterminant diffère du zéro, il existe la matrice inverse $M^{-1}(t_0, T)$, et l'équation (106) peut être résolue par rapport à $x(t_0)$:

$$x(t_0) = M^{-1}(t_0, T)\,\eta(t_0, T). \tag{11.107}$$

De la sorte, *si la matrice* $M(t_0, T)$ *est régulière, le système* (99), (100) *est complètement observable.*

Montrons maintenant que *si le système est complètement observable,* $M(t_0, T)$ *est une matrice régulière.*

Supposons que la réciproque est vraie: soit $M(t_0, T)$ une matrice singulière et

$$\langle \mu, M\mu \rangle = 0. \tag{11.108}$$

Prenons le système à l'état initial

$$x(t_0) = \mu \neq 0. \tag{11.109}$$

D'après (102), sous la condition initiale (109), la fonction $y(t)$ s'écrit

$$y(t) \equiv q(t), \tag{11.110}$$

avec

$$q(t) = C(t)\,\Phi(t, t_0)\,\mu. \tag{11.111}$$

Si l'on tient compte que

$$\langle C\Phi\mu, C\Phi\mu \rangle = (C\Phi\mu)^* C\Phi\mu = \mu^*\Phi^*C^*C\Phi\mu = \langle \mu, \Phi^*C^*C\Phi\mu \rangle,$$

on a en vertu de (104) et (108)

$$\int_{t_0}^{T} \langle q(t), q(t)\rangle\,dt = \int_{t_0}^{T} \langle C\Phi\mu, C\Phi\mu\rangle\,dt =$$

$$= \int_{t_0}^{T} \langle \mu, \Phi^*C^*C\Phi\mu\rangle\,dt = \left\langle \mu, \int_{t_0}^{T} \Phi^*C^*C\Phi\,dt\mu \right\rangle = \langle \mu, M\mu\rangle = 0. \tag{11.112}$$

Il résulte de la relation (112) que

$$q(t) \equiv 0 \qquad (t_0 \leqslant t \leqslant T). \tag{11.113}$$

Dans un système complètement observable, avec $x(t_0) \neq 0$, le vecteur fonction $y(t)$ vérifie la condition

$$y(t) \not\equiv 0 \qquad (t_0 \leqslant t \leqslant T), \tag{11.114}$$

ce qui contredit les relations (113) et (110).

Cette contradiction résulte de l'hypothèse (108) suivant laquelle la matrice $M(t_0, T)$ d'un système complètement observable est singulière. Par suite, si un système est complètement observable, la matrice $M(t_0, T)$ est régulière.

Nous avons ainsi démontré le théorème suivant.

Le système (99), (100) *est complètement observable dans l'intervalle* $[t_0, T]$ *si et seulement si*

$$M(t_0, T) = \int_{t_0}^{T} \Phi^*(t, t_0)\,C^*(t)\,C(t)\,\Phi(t, t_0)\,dt$$

est une matrice régulière.

Notons que puisque $M(t_0, T)$ est une matrice de Gram, si $\det M(t_0, T) \neq 0$ (comme c'est le cas du théorème qui vient d'être démontré), $M(t_0, T)$ est une matrice définie positive.

8. **Condition de commandabilité d'un système linéaire stationnaire dans un problème aux extrémités libres.** Considérons le système décrit par les équations vectorielles

$$\left.\begin{aligned} \frac{dx}{dt} &= Ax + Gu \\ y &= Cx, \end{aligned}\right\} \qquad (11.115)$$

où

$$\left.\begin{aligned} x &= \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \ldots & A_{nn} \end{bmatrix}, \quad G = \begin{bmatrix} G_{11} & \ldots & G_{1r} \\ \cdots & \cdots & \cdots \\ G_{n1} & \ldots & G_{nr} \end{bmatrix}, \\ u &= \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \cdots \\ y_p \end{bmatrix}, \quad C \begin{bmatrix} c_{11} & \ldots & c_{1n} \\ \cdots & \cdots & \cdots \\ c_{p1} & \ldots & c_{pn} \end{bmatrix}. \end{aligned}\right\} \qquad (11.116)$$

Les éléments des matrices A, G et C sont supposés constants. Le rang du polynôme minimal $\psi(\lambda)$ de la matrice A est égal à m.

Désignons par x_k ($k = 1, \ldots, n$) les coordonnées de phase du système, par u_l ($l = 1, \ldots, r$) les commandes qui lui sont appliquées.

y_j ($j = 1, \ldots, p$; $p \leqslant n$) désigne les combinaisons linéaires des coordonnées de phase $x_1, \ldots, x_n$ déterminées d'après (115). Le but de la commande est d'amener le vecteur $y(t)$ à l'instant $t = T$ à l'état donné

$$y(T) = K \qquad (11.117)$$

Conformément à (115) l'équation vectorielle (117) est équivalente au système d'équations scalaires

$$\sum_{k=1}^{n} c_{jk} x_k(T) = K_j \qquad (j = 1, \ldots, p). \qquad (11.118)$$

Les conditions (118) traduisent le fait que le point représentatif $(x_1, \ldots, x_n)$ doit être ramené à l'instant $t = T$ dans le plan de dimension $(n - p)$ de l'espace X de dimension n défini par les relations

$$\sum_{k=1}^{n} c_{jk} x_k = K_j \qquad (j = 1, \ldots, p). \qquad (11.119)$$

Les conditions qui rendent possible cette opération s'appellent [36] *commandabilité par rapport à y*.

Théorème. *Pour que le système* (115) *soit complètement commandable par rapport à y, il faut et il suffit que le rang de la matrice*

$$P = [CG \; CAG \; CA^2G \; \ldots \; CA^{m-1}G] \qquad (11.120)$$

soit égal à p.

D é m o n s t r a t i o n. Puisque d'après (115)

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Gu(\tau)\,d\tau, \tag{11.121}$$

la relation (117) devient

$$C\int_0^T e^{A(T-\tau)}Gu(\tau)\,d\tau = -Ce^{AT}x(0) + K. \tag{11.122}$$

En désignant par $z(t)$ le vecteur

$$z(t) = Ce^{At}x(0) - K, \tag{11.123}$$

la relation (122) peut s'écrire

$$C\int_0^T e^{A(T-\tau)}Gu(\tau)\,d\tau = -z(T). \tag{11.124}$$

D'après (10.58)

$$e^{A(T-\tau)} = \sum_{k=0}^{m-1}\alpha_k(T-\tau)A^k, \tag{11.125}$$

et c'est pourquoi

$$C\int_0^T e^{A(T-\tau)}Gu(\tau)\,d\tau = \sum_{k=0}^{m-1}CA^kG\int_0^T u(\tau)\alpha_k(T-\tau)\,d\tau. \tag{11.126}$$

Maintenant la relation (124) peut se mettre sous la forme

$$[CG\ CAG\ CA^2G\ \dots\ CA^{m-1}G] \times \begin{bmatrix} \int_0^T u(\tau)\alpha_0(T-\tau)\,d\tau \\ \int_0^T u(\tau)\alpha_1(T-\tau)\,d\tau \\ \dots\dots\dots\dots \\ \int_0^T u(\tau)\alpha_{m-1}(T-\tau)\,d\tau \end{bmatrix} = -z(T). \tag{11.127}$$

Considérons ensuite que le nombre de commandes r est choisi tel qu'il respecte la condition

$$q = mr \geqslant p. \tag{11.128}$$

La matrice P, définie par la relation (120), est une matrice rectangulaire $p \times q$:

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1q} \\ \cdot & & & \\ P_{p1} & P_{p2} & \dots & P_{pq} \end{bmatrix}. \tag{11.129}$$

Désignons par ξ le vecteur dont les éléments ξ_j $(j = 1, \ldots, q)$ sont de la forme:

$$\left.\begin{aligned} \xi_1 &= \int_0^T u_1(\tau)\,\alpha_0(T-\tau)\,d\tau, \\ \xi_2 &= \int_0^T u_2(\tau)\,\alpha_0(T-\tau)\,d\tau, \\ &\cdots\cdots\cdots\cdots\cdots \\ \xi_r &= \int_0^T u_r(\tau)\,\alpha_0(T-\tau)\,d\tau, \\ \xi_{r+1} &= \int_0^T u_1(\tau)\,\alpha_1(T-\tau)\,d\tau, \\ &\cdots\cdots\cdots\cdots\cdots \\ \xi_q &= \int_0^T u_r(\tau)\,\alpha_{m-1}(T-\tau)\,d\tau. \end{aligned}\right\} \qquad (11.130)$$

L'équation vectorielle (127) devient ainsi:

$$P\xi = -z(T). \qquad (11.131)$$

A l'équation vectorielle (131) correspond le système de p équations scalaires par rapport à q (où $q \geqslant p$) inconnues $\xi_1, \ldots, \xi_q$:

$$\left.\begin{aligned} P_{11}\xi_1 + P_{12}\xi_2 + \ldots + P_{1q}\xi_q &= -z_1(T), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ P_{p1}\xi_1 + P_{p2}\xi_2 + \ldots + P_{pq}\xi_q &= -z_p(T). \end{aligned}\right\} \qquad (11.132)$$

Si l'on note par Π_k $(k = 1, \ldots, q)$ les vecteurs

$$\Pi_k = \begin{bmatrix} P_{1k} \\ \cdots \\ P_{pk} \end{bmatrix}, \qquad (11.133)$$

on a pour le système dé'quations (132)

$$\sum_{k=1}^{q} \Pi_k \xi_k = -z(T). \qquad (11.134)$$

De cette façon le vecteur

$$z(T) = \begin{bmatrix} z_1(T) \\ \cdots \\ z_p(T) \end{bmatrix}$$

est une combinaison linéaire des vecteurs Π_k $(k = 1, \ldots, q)$, où $q \geqslant p$. Le vecteur $z(T)$ pouvant être quelconque, la condition de

résolubilité de l'équation (134) par rapport à $\xi_1, \ldots, \xi_q$ entraîne que pour rendre le système (115) complètement commandable par rapport à y il faut et il suffit qu'il existe parmi les vecteurs Π_k ($k = 1, \ldots, q$) p vecteurs linéairement indépendants, c'est-à-dire que le rang de la matrice P soit égal à p. Le théorème est démontré.

Si $q > p$, le système des variables $\xi_1, \ldots, \xi_q$ vérifiant les équations (132) *n'est pas unique*. Une fois les variables $\xi_1, \ldots, \xi_q$ trouvées, il faut définir d'après les relations (130), la loi de commande $u_l = u_l(t)$ ($0 \leqslant t \leqslant T$; $l = 1, \ldots, r$). Ce problème admet également plus d'une solution.

Montrons à titre d'exemple que parmi les commandes possibles qui observent la condition (117)

$$y(T) = K$$

il y a la commande

$$\tilde{u}(t) = -G^* e^{A^*(T-t)} C^* \Psi^{-1} [C e^{AT} x(0) - K], \qquad (11.135)$$

où Ψ désigne la matrice carrée $p \times p$

$$\Psi = \int_0^T C e^{A(T-\tau)} G G^* e^{A^*(T-\tau)} C^* d\tau. \qquad (11.136)$$

En effet, pour $u = \tilde{u}(t)$ on a

$$C e^{AT} \int_0^T e^{-A\tau} G \tilde{u}(\tau) d\tau =$$

$$= -\int_0^T C e^{A(T-\tau)} G G^* e^{A^*(T-\tau)} C^* d\tau \Psi^{-1} [C e^{AT} x(0) - K] =$$

$$= -C e^{AT} x(0) + K. \qquad (11.137)$$

Conformément à (115), (121) et (137), pour $y(T)$ il vient

$$y(T) = C e^{AT} x(0) + \int_0^T C e^{A(T-\tau)} G \tilde{u}(\tau) d\tau = K,$$

c'est-à-dire la relation (117) est vérifiée.

Nous avons supposé précédemment (135) l'existence de la matrice Ψ^{-1}. Au point 9 nous allons montrer que *si le système* (115) *est complètement commandable par rapport à* y, Ψ *est une matrice définie positive*, et, par conséquent, l'inverse Ψ^{-1} existe bien.

9. Condition de commandabilité d'un système linéaire non stationnaire dans un problème aux extrémités libres. Etudions pour un système décrit par les équations vectorielles

$$\left.\begin{aligned} \frac{dx}{dt} &= A(t)x + G(t)u, \\ y &= C(t)x, \end{aligned}\right\} \qquad (11.138)$$

où

$$x=\begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad A(t)=\begin{bmatrix} A_{11}(t) & \dots & A_{1n}(t) \\ \cdots & \cdots & \cdots \\ A_{n1}(t) & \dots & A_{nn}(t) \end{bmatrix},$$

$$G(t)=\begin{bmatrix} G_{11}(t) & \dots & G_{1r}(t) \\ \cdots & \cdots & \cdots \\ G_{n1}(t) & \dots & G_{nr}(t) \end{bmatrix},$$

$$u=\begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad y=\begin{bmatrix} y_1 \\ \cdots \\ y_p \end{bmatrix}, \quad C=\begin{bmatrix} c_{11}(t) & \dots & c_{1n}(t) \\ \cdots & \cdots & \cdots \\ c_{p1}(t) & \dots & c_{pn}(t) \end{bmatrix}, \quad (11.139)$$

le problème qui consiste à ramener le vecteur $y(t)$ à l'instant $t = t_1$ à l'état donné

$$y(t_1) = K. \quad (11.140)$$

Ce problème a été étudié au point 8 pour les systèmes linéaires stationnaires où la possibilité d'observer la relation (140) a été nommée commandabilité par rapport à y.

Théorème. *Soit* $\Phi(t, \tau) = \theta(t)\,\theta^{-1}(\tau)$, *où* $\theta(t)$ *est la matrice fondamentale des solutions du système décrit par l'équation différentielle vectorielle homogène*

$$\frac{dx}{dt} = A(t)\,x, \quad (11.141)$$

et $V(t_0, t_1)$ *la matrice*

$$V(t_0, t_1) = \int_{t_0}^{t_1} C(t_1)\,\Phi(t_1, \tau)\,G(\tau)\,G^*(\tau)\,\Phi^*(t_1, \tau)\,C^*(t_1)\,d\tau. \quad (11.142)$$

Le système linéaire non stationnaire (138) *est complètement commandable par rapport à* y *dans l'intervalle de temps* $t_0 \leqslant t \leqslant t_1$ *si et seulement si la matrice* $V(t_0, t_1)$ *déterminée par l'expression* (142) *est une matrice définie positive.*

Démonstration.

1°. Par condition $V(t_0, t_1)$ est une matrice définie positive. Il en résulte que c'est une matrice régulière, c'est-à-dire $\det V(t_0, t_1) \neq 0$. C'est pourquoi la matrice inverse $V^{-1}(t_0, t_1)$ existe bien et on peut choisir le vecteur des commandes $u(t)$ tel que

$$\tilde{u}(t) = -G^*(t)\,\Phi^*(t_1, t)\,C^*(t_1)\,V^{-1}(t_0, t_1)[C(t_1)\,\Phi(t_1, t_0)\,x(t_0) - K]. \quad (11.143)$$

D'après (138) l'état du système à l'instant $t = t_1$ s'écrit

$$x(t_1) = \Phi(t_1, t_0)\,x(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \tau)\,G(\tau)\,u(\tau)\,d\tau. \quad (11.144)$$

Avec la commande $u = \tilde{u}(t)$ on a, grâce à (138) et (144),

$$y(t_1) = C(t_1)\,x(t_1) = C(t_1)\,\Phi(t_1, t_0)\,x(t_0) + C(t_1)\int_{t_0}^{t_1}\Phi(t_1, \tau)\,G(\tau)\,\tilde{u}(\tau)\,d\tau. \tag{11.145}$$

Comme

$$C(t_1)\int_{t_0}^{t_1}\Phi(t_1, \tau)\,G(\tau)\,\tilde{u}(\tau)\,d\tau =$$

$$= -\int_{t_0}^{t_1} C(t_1)\,\Phi(t_1, \tau)\,G(\tau)\,G^*(\tau)\,\Phi^*(t_1, \tau)\,C(t_1)\,d\tau\,V^{-1}(t_0, t_1)\,\times$$

$$\times\,[C(t_1)\,\Phi(t_1, t_0)\,x(t_0) - K] = -C(t_1)\,\Phi(t_1, t_0)\,x(t_0) + K, \tag{11.146}$$

l'expression (145) devient

$$y(t_1) = K.$$

Ainsi, la commande $\tilde{u}(t)$, déterminée par l'expression (143), assure l'observation de la relation (140).

Notons que la commande $\tilde{u}(t)$ donnée par l'expression (143), qui, comme nous venons de le démontrer, amène le vecteur $y(t)$ à l'état $y(t_1) = K$, *n'est pas unique*. En effet, la commande

$$u(t) = \tilde{u}(t) + v(t), \tag{11.147}$$

où $v(t)$ est un vecteur fonction quelconque de dimension r vérifiant la condition

$$C(t_1)\int_{t_0}^{t_1}\Phi(t_1, \tau)\,G(\tau)\,v(\tau)\,d\tau = 0 \tag{11.148}$$

amène également d'après (145) à l'instant $t = t_1$ le vecteur $y(t)$ à l'état $y(t_1) = K$. La première partie du théorème est démontrée.

2°. La démonstration de la deuxième partie du théorème selon laquelle si le système (138) est complètement commandable par rapport à y, la matrice $V(t_0, t_1)$ est définie positive, reprend fidèlement la démonstration de l'affirmation analogue relative à la matrice $W(t_0, t_1)$ formulée au point 6, section 2°, p. 189.

CHAPITRE 4

MÉTHODE DE PROGRAMMATION DYNAMIQUE

§ 12. Commande optimale des systèmes à ressources restreintes

1. Problèmes variationnels dans la théorie de la commande. Lors de l'étude des problèmes de commandabilité au § 11 nous n'avons pas imposé aux commandes des contraintes a priori. Nous avons également considéré comme admissible (dans le choix de telle ou telle commande) toute loi de variation dans le temps des coordonnées de phase du système.

Or, en pratique les ressources d'un système commandé sont limitées. Ainsi, par exemple, des contraintes sont possibles limitant le module de la commande, la puissance de sa source, etc. Il se peut également que pour telle ou telle raison les écarts admissibles de certaines coordonnées de phase du système soient limitées, etc.

Le choix de la loi de commande (c'est-à-dire de la loi de variation des commandes), compte tenu des contraintes citées, est guidé par le but imposé. Dans le cas général, on peut admettre que le but de la commande est d'obtenir l'extrémum d'une fonctionnelle qui caractérise le critère d'optimalité du système

$$Q(x, z, u, t) = \min. \tag{12.1}$$

Ici x est le vecteur de dimension n dont les éléments $x_1, \ldots, x_n$ sont les coordonnées de phase; z, le vecteur de dimension m dont les éléments $z_1, \ldots, z_m$ sont les signaux d'entrée à reproduire par le système ou à transformer de la façon donnée; u, le vecteur de dimension r dont les éléments $u_1, \ldots, u_r$ sont les commandes; t, le temps.

La fonctionnelle Q est un nombre dépendant de la forme des vecteurs fonctions $x(t)$, $z(t)$ et $u(t)$. Ainsi, par exemple, la fonctionnelle Q peut s'écrire

$$Q = \sum_{i=1}^{m} \int_0^T [x_{\sigma_i}(t) - z_i(t)]^2 \, dt, \tag{12.2}$$

où T est une grandeur fixée. La fonctionnelle Q qui, d'après (2), dépend de la forme des fonctions $x_{\sigma_i}(t)$, $z_i(t)$ $(i = 1, \ldots, m)$ dans l'intervalle de temps $0 < t < T$, est la mesure de la qualité de reproduction par le système du signal d'entrée vectoriel $z(t)$.

. Dans de nombreux problèmes on demande que la trajectoire du point représentatif, qui passe par deux points donnés x (0) et x (T) de l'espace de phase, où T est une grandeur fixée (c'est-à-dire que la trajectoire de phase qui amène le système de l'état initial x (0) à un état donné x (T), où T est un instant donné à l'avance), minimise ou maximise la fonctionnelle

$$Q=\int_0^T G(x(t),\ u(t),\ t)\,dt, \tag{12.3}$$

avec G, une fonction scalaire bornée des variables $x_1, \ldots, x_n, u_1, u_2, \ldots, u_r, t$.

La forme de la fonction $G(x, u, t)$ se définit pour chaque problème concret. Ainsi, si l'on donne l'état final x (T) sans fixer à l'avance l'instant final T et si la fonction G s'écrit $G \equiv 1$, on tombe sur $Q = T$, et la condition (1) $Q = \min$ devient $T = \min$. On aboutit ainsi à un problème de réponse en temps minimal dans lequel il faut établir la loi de commande $u = u(t)$ vérifiant les contraintes imposées à u telle qu'en un temps minimal T le système serait amené de l'état initial x (0) à l'état final prévu x (T).

Ce qui précède caractérise la position des problèmes de commande optimale. Ce sont des problèmes variationnels qui consistent à rechercher l'extrémum lié des fonctionnelles dont la forme est déterminée par le critère d'optimalité adopté.

2. Méthode de programmation dynamique de R. Bellman. Principe d'optimalité. Considérons un système de commande décrit par le système d'équations différentielles scalaires

$$\frac{dx_j}{dt}=f_j(x_1, \ldots, x_n, u_1, \ldots, u_r) \quad (j=1, \ldots, n). \tag{12.4}$$

Ici $x_1, \ldots, x_n$ sont les coordonnées de phase du système; $u_1, \ldots, u_r$, les commandes.

En introduisant les vecteurs

$$\left.\begin{array}{c} x=\begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u=\begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \\ f(x, u)=\begin{bmatrix} f_1(x_1, \ldots, x_n, u_1, \ldots, u_r) \\ \cdots\cdots\cdots\cdots \\ f_n(x_1, \ldots, x_n, u_1, \ldots, u_r) \end{bmatrix}, \end{array}\right\} \tag{12.5}$$

on peut remplacer le système (4) par l'équation différentielle vectorielle:

$$\frac{dx}{dt}=f(x, u). \tag{12.6}$$

En posant que les commandes $u_1, \ldots, u_r$ sont soumises à certaines contraintes, demandons que leur choix soit guidé par l'observation de la condition

$$u \in \Omega, \tag{12.7}$$

où Ω est un domaine de l'espace $(u_1, \ldots, u_r)$, défini par la forme des contraintes imposées.

Supposons que le *but de la commande* consiste à minimiser la fonctionnelle

$$Q = \int_0^T G(x(t), u(t))\, dt, \tag{12.8}$$

où G est une fonction scalaire bornée des variables $x_1, \ldots, x_n$, $u_1, \ldots, u_r$, et T, une grandeur fixée donnée.

La méthode de programmation dynamique est basée sur le *principe d'optimalité* énoncé par R. Bellman [8]. Ce principe est vrai pour les systèmes dont le mouvement ultérieur est complètement défini par leur état à un instant courant quelconque. Tels sont, par exemple, les systèmes régis par les équations différentielles (4), où par état on entend la position du système dans l'espace de phase, les systèmes décrits par les équations aux différences finies à variable discrète, etc. Voici la formulation du principe d'optimalité de R. Bellman.

Un comportement est optimal si, quels que soient l'état initial et la solution à l'instant initial, les solutions ultérieures constituent un comportement optimal par rapport au résultat de la première solution.

Cette formulation du principe d'optimalité (qualifié par Bellman d'intuitif) concerne une classe très générale des systèmes. Par « comportement » des systèmes commandés décrits par les équations différentielles (4) on entend leur mouvement et le terme « solution » se rapporte au choix de la loi régissant la variation des commandes dans le temps.

Si on inclut dans la notion de l'état d'un système à un instant donné t^* tout ce qui précédait la variation des coordonnées de phase du système dans l'intervalle de postaction $t^* - \tau < t < t^*$, le principe d'optimalité qui vient d'être énoncé sera également vrai pour les systèmes à postaction, c'est-à-dire régis par des équations aux différences.

Pour les systèmes décrits par les équations différentielles (4) le principe d'optimalité coïncide avec le fait bien connu qu'une partie de trajectoire optimale est encore une trajectoire optimale.

La figure 12.1 représente à titre d'exemple [85] la trajectoire optimale du système (4) qui passe par le point donné $x(0)$, c'est-à-dire la trajectoire qui, sous la condition (7), minimise la fonctionnelle (8) où la grandeur T est supposée fixée. On suppose également que la valeur $x(T)$ est inconnue à l'avance. Le point $x(t_1)$ divise la

trajectoire étudiée en deux arcs *1* et *2*. L'arc *2* est associé à la fonctionnelle

$$Q_2 = \int_{t_1}^{T} G(x(t), u(t))\, dt. \tag{12.9}$$

L'arc *2* peut être considéré également comme une trajectoire indépendante qui sera optimale si elle minimise la fonctionnelle (9).

Le principe d'optimalité dit que *l'arc 2 de la trajectoire optimale 1-2 est lui-même une trajectoire optimale du système* (4), dont l'état pour $t = t_1$ est $x(t_1)$.

Si l'on admet le contraire, il existe (fig. 12.1) une autre trajectoire *2'* qui donne à la fonctionnelle (9) une valeur plus petite par rapport à la trajectoire *2*. Mais la trajectoire optimale dans l'intervalle de temps (0, *T*) sera *1-2'*, et non pas *1-2*. On aboutit ainsi à une contradiction avec les données initiales suivant lesquelles la trajectoire *1-2* est optimale. Cette contradiction prouve que dans l'intervalle (t_1, *T*) l'arc *2* de la trajectoire optimale *1-2* est encore une trajectoire optimale du système (4).

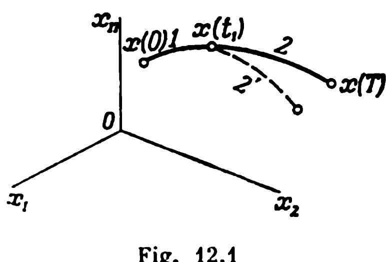


Fig. 12.1

Notons maintenant que le principe d'optimalité concerne le mouvement du système qui succède à l'état donné. En général, pour un état antérieur à l'état donné il peut ne pas être vérifié.

Ainsi, par exemple, si l'on ne se donne que l'état initial du système $x(0)$, l'arc *1* de la trajectoire optimale *1-2* peut ne pas être lui-même une trajectoire optimale, c'est-à-dire il se peut qu'il ne minimise pas la fonctionnelle

$$Q_1 = \int_{0}^{t_1} G(x(t), u(t))\, dt. \tag{12.10}$$

Ce n'est que lorsque le point final $x(t_1)$ de l'arc *1* est donné que cet arc sera par lui-même une trajectoire optimale.

Donc, dans le cas de systèmes de commande, le principe d'optimalité dit que *le choix de la commande optimale n'est défini que par l'état du système à l'instant courant.*

Cette affirmation permet d'obtenir les équations fonctionnelles données ci-dessous qui déterminent la loi de variation des commandes dans le problème de la commande optimale.

La méthode de Bellman relative à la détermination de la commande optimale est intimement liée au problème variationnel de

propagation de l'excitation [24] et conduit aux équations aux dérivées partielles de Jacobi-Hamilton.

Pour les systèmes discrets, la méthode de Bellman permet de déterminer les commandes en plusieurs étapes.

§ 13. Application de la programmation dynamique aux systèmes discrets

1. Relation de récurrence de Bellman. De nombreux systèmes de commande sont décrits par des équations aux différences finies. Ces systèmes portent généralement le nom de systèmes discrets. On range dans cette catégorie les systèmes impulsionnels, les systèmes munis de calculateurs digitaux, etc.

Les systèmes décrits par des équations différentielles s'appellent en ce sens systèmes continus.

A tout système d'équations différentielles on peut faire correspondre un système équivalent d'équations aux différences finies à variable continue [77]. A cet effet, il faut intégrer le système d'équations différentielles donné dans l'intervalle de temps fini $t_0 + \varepsilon \leqslant t < t_0 + \varepsilon + \tau$, où τ est l'intervalle de discontinuité fixé et ε (où $0 < \varepsilon < \tau$), le paramètre associé à un point situé à l'intérieur de l'intervalle de discontinuité. Les coefficients du système d'équations aux différences finies dépendent de ε et la solution de ce système d'équations définit, à tout instant t, l'état du système considéré.

Le système d'équations aux différences finies ainsi obtenu constitue des équations fonctionnelles exactes qui associent entre eux les états du système aux instants séparés par l'intervalle de discontinuité τ.

Si l'on se borne à l'étude de l'état du système aux instants discrets séparés par des intervalles de discontinuité τ, on peut alors fixer la valeur du paramètre ε et obtenir ainsi un système d'équations aux différences finies à variable discrète. Sa solution détermine exactement l'état du système à des instants discrets séparés par la grandeur $\tau =$ const. La position de ces instants dans l'intervalle de discontinuité est définie par le choix du paramètre ε.

La construction des équations aux différences finies mentionnées ci-dessus qui décrivent exactement le système de commande à action continue peut s'avérer très longue dans le cas des problèmes suffisamment complexes.

Les équations aux différences finies *approchées* peuvent s'obtenir de la façon suivante.

Supposons que le système de commande est régi par l'équation différentielle scalaire

$$\frac{d\tilde{x}}{dt} = f_1(\tilde{x}(t), \tilde{u}(t)), \tag{13.1}$$

où $\tilde{x}$ est la coordonnée du système et $\tilde{u}$, la commande qui doit vérifier la contrainte

$$\tilde{u} \in \Omega. \tag{13.2}$$

L'état initial du système est donné

$$[\tilde{x}(t)]_{t=0} = \tilde{x}(0),$$

et il faut trouver la commande $\tilde{u} = \tilde{u}(t)$ minimisant la fonctionnelle

$$Q = \int_0^T G_1(\tilde{x}(t), \tilde{u}(t))\,dt + \varphi(\tilde{x}(T)), \tag{13.3}$$

avec T, une certaine grandeur fixée. Considérons que

$$T = N\tau, \tag{13.4}$$

où N est un nombre entier et $\tau = \text{const}$, une grandeur suffisamment petite que nous allons adopter comme intervalle de discontinuité. Bornons-nous ici à l'étude de l'état du système aux instants discrets

$$t = \vartheta\tau \qquad (\vartheta = 0, 1, 2, \ldots, N). \tag{13.5}$$

Comme

$$\frac{d\tilde{x}}{dt} \approx \frac{\tilde{x}(t+\tau) - \tilde{x}(t)}{\tau} = \frac{\tilde{x}(\vartheta\tau+\tau) - \tilde{x}(\vartheta\tau)}{\tau}, \tag{13.6}$$

l'équation différentielle initiale (1) peut être remplacée par la relation approchée suivante:

$$\tilde{x}(\vartheta\tau + \tau) - \tilde{x}(\vartheta\tau) = \tau f_1(\tilde{x}(\vartheta\tau), \tilde{u}(\vartheta\tau)). \tag{13.7}$$

Le premier membre de (7) est la différence première de la fonction $\tilde{x}(t)$; la relation (7) représente donc une équation aux différences finies par rapport à la fonction scalaire inconnue $\tilde{x}(\vartheta\tau)$.

La variable des fonctions $\tilde{x}$ et $\tilde{u}$ du second membre de (7) étant discrète, (7) est une équation aux différences finies à variable discrète et sa solution ne définit l'état du système qu'aux instants discrets $t = \vartheta\tau$ ($\vartheta = 1, 2, \ldots$).

La fonctionnelle (3) est approximée par la somme intégrale

$$Q = \tau \sum_{\vartheta=0}^{N-1} G_1(\tilde{x}(\vartheta\tau), \tilde{u}(\vartheta\tau)) + \varphi(\tilde{x}(N\tau)). \tag{13.8}$$

En introduisant les notations

$$x(\vartheta) = \tilde{x}(\vartheta\tau), \qquad u(\vartheta) = \tilde{u}(\vartheta\tau), \tag{13.9}$$

$$\left.\begin{aligned} f(x(\vartheta), u(\vartheta)) &= \tau f_1(x(\vartheta), u(\vartheta)), \\ G(x(\vartheta), u(\vartheta)) &= \tau G_1(x(\vartheta), u(\vartheta)), \end{aligned}\right\} \tag{13.10}$$

on aboutit, d'après (7), (2) et (8), au problème suivant.

Trouver pour un système décrit par l'équation scalaire aux différences finies à variable discrète

$$x(\vartheta + 1) = x(\vartheta) + f(x(\vartheta),\ u(\vartheta)), \tag{13.11}$$

dont l'état initial est donné

$$[x(\vartheta)]_{\vartheta=0} = x(0), \tag{13.12}$$

la commande $u = u(\vartheta)$ qui satisfait aux contraintes

$$u \in \Omega \tag{13.13}$$

et minimise la somme

$$Q = \sum_{\vartheta=0}^{N-1} G(x(\vartheta),\ u(\vartheta)) + \varphi(x(N)), \tag{13.14}$$

avec N, une grandeur fixée.

Dans ce qui suit nous allons fonder l'étude des systèmes discrets sur les équations de la forme (11)-(14) (en posant, dans le cas général, que x, u, f sont des vecteurs de dimension correspondante), sans plus parler du mode de leur obtention pour un système concret mais en considérant quand même que ces équations sont exactes (du fait que, dans le cas général, on peut obtenir des équations aux différences exactes de la forme (11)).

L'erreur des solutions fournies par l'approximation (6) (donnée à titre d'illustration) des équations aux différences peut devenir importante et chaque fois le choix d'une telle approximation doit être justifié. Toutefois cette question exige des études spéciales et nous n'allons pas nous y attarder.

Le problème posé ci-dessus décrit par les équations (11)-(14) consiste à chercher un ensemble des variables $u(0)$, $u(1)$, $u(2)$, ..., $u(N-1)$ minimisant la somme (14) sous les conditions (11), (12) et (13).

Pour le résoudre, commençons par le dernier intervalle de temps $(N-1)\tau \leqslant t < N\tau$, en supposant connu l'état $x(N-1)$. D'après le principe d'optimalité, dans l'intervalle de temps $(N-1)\tau \leqslant$ $\leqslant t < N\tau$ la commande u doit être choisie compte tenu de la contrainte (13) de façon à minimiser la somme partielle

$$Q_{N-1} = G(x(N-1),\ u(N-1)) + \varphi(x(N)), \tag{13.15}$$

qui correspond à cet intervalle de temps. Etant donné que, d'après (11),

$$x(N) = x(N-1) + f(x(N-1),\ u(N-1)), \tag{13.16}$$

l'expression (15) devient

$$Q_{N-1} = G(x(N-1),\ u(N-1)) + \varphi[x(N-1) + \\ + f(x(N-1),\ u(N-1))]. \tag{13.17}$$

Puisque l'état $x(N-1)$ est supposé connu, l'expression (17) ne dépend que d'*une seule* grandeur inconnue $u(N-1)$ qu'il faut trouver en tenant compte de la contrainte (13) en minimisant la grandeur

Q_{N-1}. Désignons par $u^*(N-1)$ la valeur optimale obtenue de $u(N-1)$. La valeur minimale Q_{N-1} qu'on obtient alors est notée S_{N-1}:

$$\min_{u(N-1)\in\Omega} Q_{N-1} = S_{N-1} = S_{N-1}(x(N-1)). \tag{13.18}$$

(18) montre d'une façon explicite que la valeur minimale Q_{N-1} dépend de l'état du système $x(N-1)$.

Conformément à (17) la relation (18) devient

$$S_{N-1}(x(N-1) = \min_{u(N-1)\in\Omega} Q_{N-1} =$$
$$= \min_{u(N-1)\in\Omega} [G(x(N-1),\ u(N-1)) + \varphi(x(N))] =$$
$$= \min_{u(N-1)\in\Omega} \{G(x(N-1),\ u(N-1)) + \varphi[x(N-1) +$$
$$+ f(x(N-1),\ u(N-1))]\}. \tag{13.19}$$

Considérons maintenant l'intervalle de temps $(N-2)\tau \leqslant t < N\tau$, composé du dernier et de l'avant-dernier intervalles. D'après (14), à cet intervalle correspond la somme partielle

$$Q_{N-2} = G(x(N-2), u(N-2)) + G(x(N-1), u(N-1)) +$$
$$+ \varphi(x(N)) \tag{13.20}$$

ou, en vertu de (15),

$$Q_{N-2} = G(x(N-2),\ u(N-2)) + Q_{N-1}. \tag{13.21}$$

Supposons que l'état $x(N-2)$ est connu. Le principe d'optimalité entraîne que, dans l'intervalle $[(N-2)\tau,\ N\tau)$, ce ne sont que l'état $x(N-2)$ et le but de la commande (minimisation de Q_{N-2}) qui déterminent la commande optimale.

Cherchons maintenant S_{N-2}, c'est-à-dire le minimum de Q_{N-2} par rapport à $u(N-2)$ et $u(N-1)$. Retenons que le minimum de la somme partielle Q_{N-1} par rapport à $u(N-1)$ est déjà trouvé et que ce minimum

$$S_{N-1} = S_{N-1}(x(N-1))$$

figure sous la forme d'une fonction de $x(N-1)$. Comme d'après (11)

$$x(N-1) = x(N-2) + f(x(N-2),\ u(N-2)), \tag{13.22}$$

S_{N-1} s'écrit

$$S_{N-1}(x(N-1)) = S_{N-1}[x(N-2) + f(x(N-2),\ u(N-2))]. \tag{13.23}$$

Puisque le premier terme du second membre de (21) ne dépend pas de $u(N-1)$, on a

$$S_{N-2}(x(N-2)) = \min_{\substack{u(N-2)\in\Omega \\ u(N-1)\in\Omega}} Q_{N-2} =$$

$$= \min_{u(N-2)\in\Omega} [G(x(N-2),\ u(N-2)) + S_{N-1}(x(N-1))] =$$

$$= \min_{u(N-2)\in\Omega} \{G(x(N-2),\ u(N-2)) + S_{N-1}[x(N-2) + $$

$$+ f(x(N-2),\ u(N-2))]\}. \quad (13.24)$$

Ainsi, l'expression (24) est ramenée à une forme qui n'impose la minimisation que par rapport à une seule variable, et notamment par rapport à $u(N-2)$. La valeur $u(N-2)$ ainsi obtenue est précisément la valeur optimale recherchée que nous allons noter $u^*(N-2)$.

L'expression (24) montre que le choix de $u(N-2)$ en n'imposant la minimisation que du premier terme de (21), c'est-à-dire en n'appliquant le principe d'optimalité qu'à l'intervalle $(N-2)\tau \leqslant t < (N-1)\tau$, serait incorrect, ce que nous avons déjà dit en décrivant le principe d'optimalité.

Considérons maintenant l'intervalle $(N-3)\tau \leqslant t < N\tau$, composé de trois derniers intervalles. D'après (14), à cet intervalle correspond la somme partielle

$$Q_{N-3} = G(x(N-3),\ u(N-3)) + G(x(N-2),\ u(N-2)) +$$

$$+ G(x(N-1),\ u(N-1)) + \varphi(x(N)) \quad (13.25)$$

ou, en vertu de (13.20),

$$Q_{N-3} = G(x(N-3),\ u(N-3)) + Q_{N-2}. \quad (13.26)$$

L'état $x(N-3)$ est supposé connu. D'après le principe d'optimalité, dans l'intervalle $[(N-3)\tau,\ N\tau)$, seuls l'état $x(N-3)$ et le but de la commande (minimisation de Q_{N-3}) déterminent la commande optimale.

Cherchons maintenant la grandeur S_{N-3}, c'est-à-dire le minimum de Q_{N-3} par rapport à $u(N-3)$, $u(N-2)$ et $u(N-1)$. Retenons que le minimum de la somme partielle Q_{N-2} par rapport à $u(N-2)$ et $u(N-1)$ est déjà trouvé (24) et que ce minimum

$$S_{N-2} = S_{N-2}(x(N-2))$$

est présenté sous la forme d'une fonction de $x(N-2)$. Comme d'après (11)

$$x(N-2) = x(N-3) + f(x(N-3),\ u(N-3)), \quad (13.27)$$

S_{N-2} s'écrit

$$S_{N-2}(x(N-2)) = S_{N-2}[x(N-3) + f(x(N-3),\ u(N-3))]. \quad (13.28)$$

Comme le premier terme du second membre de (26) ne dépend pas de $u\ (N-2)$ et $u\ (N-1)$, on a

$$S_{N-3}(x(N-3)) = \min_{\substack{u(N-3)\in\Omega\\ u(N-2)\in\Omega\\ u(N-1)\in\Omega}} Q_{N-3} =$$

$$= \min_{u(N-3)\in\Omega} [G(x(N-3),\ u(N-3)) + S_{N-2}(x(N-2))] =$$

$$= \min_{u(N-3)\in\Omega} \{G(x(N-3),\ u(N-3)) +$$

$$+ S_{N-2}[x(N-3) + f(x(N-3),\ u(N-3))]\}. \quad (13.29)$$

L'expression (29) est réduite à une forme qui n'impose la minimisation que par rapport à une seule variable, et notamment, par rapport à $u\ (N-3)$. La valeur de $u\ (N-3)$ ainsi trouvée est la valeur optimale recherchée que nous allons désigner par $u^*\ (N-3)$.

En considérant l'intervalle $(N-k)\ \tau \leqslant t < N\tau$ $(k = 2, 3, \ldots$ $\ldots, N)$, on obtient sans peine la formule générale (de récurrence). D'une façon analogue à (24) et (29), on a

$$S_{N-k}(x(N-k)) = \min_{\substack{u(N-k)\in\Omega\\ u(N-k+1)\in\Omega\\ \cdots\cdots\\ u(N-1)\in\Omega}} Q_{N-k} = \min_{u(N-k)\in\Omega} \{G(x(N-k),\ u(N-k)) +$$

$$+ S_{N-k+1}[x(N-k) + f(x(N-k),\ u(N-k))]\} \qquad (k = 2, 3, \ldots, N). \quad (13.30)$$

L'expression (30) est réduite à une forme qui n'impose la minimisation que par rapport à une seule variable, et notamment, par rapport à $u\ (N-k)$. La valeur de $u\ (N-k)$ ainsi obtenue est la valeur optimale recherchée que nous allons noter $u^*\ (N-k)$.

Ainsi, (19) et (30) donnent

$$\left.\begin{array}{c} u^*(N-1) = \hat{u}(x(N-1)), \\ u^*(N-2) = \hat{u}(x(N-2)), \\ \cdots\cdots\cdots\cdots\cdots \\ u^*(1) = \hat{u}(x(1)), \\ u^*(0) = \hat{u}(x(0)). \end{array}\right\} \quad (13.31)$$

Les expressions (31) montrent clairement que les valeurs $u^*\ (N-j)$ obtenues par étude des intervalles de temps partiels $[(N-j)\tau,$ $N\tau)$ $(j = 1, 2, \ldots, N)$ sont exprimées sous la forme de fonctions de l'état du système $x\ (N-j)$ au début de ces intervalles.

L'état x (0) étant donné, on trouve la valeur u^* (0). Puis d'après (11) on cherche l'état x (1) et détermine la valeur u^* (1). Ensuite (11) permet d'obtenir l'état x (2) et calculer la valeur u^* (2), etc.

Ainsi, la méthode exposée permet de bien définir la commande optimale dans l'intervalle de temps donné $[0, N\tau)$.

Dans l'algorithme de programmation dynamique du système (11) établi dans ce qui précède, le rôle essentiel revient à l'hypothèse que l'instant final du fonctionnement du système $T = N\tau$ est fixé. La construction de l'algorithme de programmation dynamique pour une valeur de T non fixée à l'avance, comme c'est le cas, par exemple, du problème de la réponse en temps minimal, est exposée dans ce qui suit.

2. Systèmes discrets multivariables. L'algorithme de programmation dynamique du système (11) peut également être appliqué à un système régi par une équation vectorielle aux différences finies ou un système d'équations scalaires aux différences finies

$$x_j(\vartheta + 1) = x_j(\vartheta) + f_j(x_1(\vartheta), \ldots, x_n(\vartheta), u_1(\vartheta), \ldots, u_r(\vartheta)) \quad (13.32)$$
$$(j = 1, \ldots, n).$$

Si l'on introduit les vecteurs

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u) = \begin{bmatrix} f_1(x, u) \\ \cdots\cdot \\ f_n(x, u) \end{bmatrix}, \quad (13.33)$$

où, pour condenser l'exposé, on désigne

$$f_j(x, u) = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r),$$

on peut remplacer le système d'équations aux différences (32) par une équation aux différences vectorielle

$$x(\vartheta + 1) = x(\vartheta) + f(x(\vartheta), u(\vartheta)). \quad (13.34)$$

Les fonctions Q, G et φ de l'expression (14) sont ici des fonctions scalaires des variables vectorielles $x(\vartheta)$, $u(\vartheta)$ $(\vartheta = 0, 1, \ldots, N - 1)$ et $x(N)$.

On voit aisément que si seulement par x, u et f on entend les vecteurs (33), les relations (19) et (30) gardent leur forme.

La différence essentielle consiste pourtant dans le fait que l'expression entre accolades du second membre de la relation (19) est maintenant une fonction de r variables scalaires $u_1(N - 1)$, $u_2(N - 1)$, $\ldots$, $u_r(N - 1)$ et sa minimisation s'effectue par rapport à toutes ces variables. D'une façon analogue, dans la relation (30) la minimisation doit se faire par rapport à r variables $u_1(N - k)$, $u_2(N - k)$, $\ldots$, $u_r(N - k)$.

La fonction scalaire

$$S_{N-j} = S_{N-j}(x(N - j)) \qquad (j = 1, \ldots, N)$$

est maintenant une fonction de la variable vectorielle $x(N-j)$. La commande optimale calculée à chaque étape (vecteur $u^*(N-j)$), est de même que dans (31) la fonction de l'état du système

$$u^*(N-j) = \hat{u}(x(N-j)) \qquad (j = 1, \ldots, N). \qquad (13.35)$$

La fonction $\hat{u}(x(N-j))$ est une fonction vectorielle de la variable vectorielle $x(N-j)$.

§ 14. Application de la programmation dynamique aux systèmes à action continue

1. Problème à temps fixé et à extrémité de trajectoire libre. Considérons le système commandé décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = f(x, u, t), \qquad (14.1)$$

où x, u et f sont les vecteurs de la forme

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u, t) = \begin{bmatrix} f_1(x, u, t) \\ \cdots\cdots \\ f_n(x, u, t) \end{bmatrix}. \qquad (14.2)$$

L'état initial du système est donné

$$[x(t)]_{t=0} = x(0). \qquad (14.3)$$

Chercher la commande $u = u(t)$ qui vérifie les contraintes

$$u \in \Omega \qquad (14.4)$$

et minimise la fonctionnelle

$$Q = \int_0^T G(x(t), u(t), t)\, dt, \qquad (14.5)$$

où T est une certaine grandeur fixée.

Soit la trajectoire optimale du système (1) (fig. 14.1) qui passe par le point donné $x(0)$, c'est-à-dire une trajectoire minimisant sous la condition (4) la fonctionnelle (5). Dans le problème considéré, la valeur $x(T)$ n'est pas donnée à l'avance.

Désignons par $S(x(0), 0)$ la valeur minimale de la fonctionnelle Q associée à la trajectoire optimale.

La figure 14.1 indique les positions du point représentatif sur la trajectoire optimale aux instants t et $t' = t + \Delta t$. L'état du système à l'instant t est $x(t)$. L'état du système à l'instant $t' = t + \Delta t$ s'écrit

$$x(t + \Delta t) = x(t') = x'. \qquad (14.6)$$

Selon le principe d'optimalité l'arc de la trajectoire optimale depuis le point x (t) jusqu'au point x (T) est lui-même une trajectoire optimale qui minimise la fonctionnelle

$$Q_t = \int_t^T G(x, u, \nu)\, d\nu. \qquad (14.7)$$

Désignons cette valeur minimale de (7) par S $(x$ (t), $t)$, ou, pour abréger l'écriture, par S (x, t).

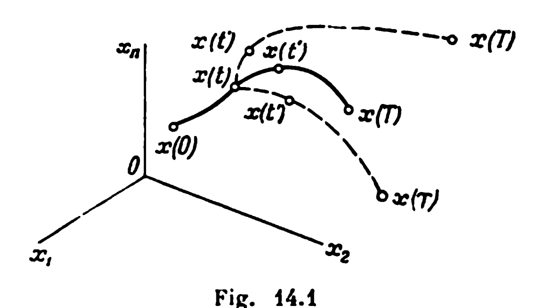


Fig. 14.1

D'une façon analogue, l'arc de la trajectoire optimale depuis le point x (t') (où $t' = t + \Delta t$) jusqu'au point x (T) est également lui-même une trajectoire optimale minimisant la fonctionnelle

$$Q_{t'} = \int_{t'}^T G(x, u, \nu)\, d\nu \qquad (t' = t + \Delta t). \qquad (14.8)$$

Introduisons les notations de la valeur minimale de la fonctionnelle (8)

$$S\ (x\ (t + \Delta t),\ t + \Delta t) = S\ (x',\ t'). \qquad (14.9)$$

Suivant la définition adoptée plus haut,

$$S(x, t) = \min_{u(\nu)\in\Omega} \int_t^T G(x(\nu),\ u(\nu),\ \nu)\, d\nu. \qquad (14.10)$$

Compte tenu de la petitesse de Δt, l'intégrale du second membre de l'expression (10) peut être mise sous la forme

$$\int_t^T G(x(\nu),\ u(\nu),\ \nu)\, d\nu = \int_t^{t'=t+\Delta t} G(x(\nu), u(\nu), \nu)\, d\nu +$$
$$+ \int_{t'=t+\Delta t}^T G(x(\nu), u(\nu), \nu)\, d\nu = G(x(t), u(t), t)\, \Delta t +$$
$$+ o_1(\Delta t) + \int_{t'}^T G(x(\nu), u(\nu), \nu)\, d\nu, \qquad (14.11)$$

où l'on suppose que la fonction u (ν) est continue dans l'intervalle semi-ouvert $[t,\ t+\Delta t)$. L'expression (10) devient

$$S(x,t)=\min_{u(t)\in\Omega}\Big[G(x(t),u(t),t)\,\Delta t+ \\ +\min_{u(\nu)\in\Omega}\int_{t'}^{T}G(x(\nu),u(\nu),\nu)\,d\nu\Big]+o_1(\Delta t). \qquad (14.12)$$

Le premier terme entre crochets de (12) ne dépend que de la valeur de la commande u à l'instant t, c'est-à-dire de $u(t)$. Le deuxième terme entre crochets doit être minimisé par rapport à l'ensemble des commandes admissibles $u(\nu)$ dans l'intervalle de temps $t'<\nu<T$. Cette valeur minimale, que nous allons noter $S(x',t')$, est à son tour fonction de l'état du système $x'=x(t')$. Quant à ce dernier, il dépend de la commande u dans l'intervalle $(t,\ t+\Delta t)$ ou, à $o_1(\Delta t)$ près, de la valeur $u(t)$ à l'instant t. En effet, en choisissant les différentes valeurs de la commande u à l'instant t on obtient un faisceau de trajectoires issu du point $x(t)$ de la trajectoire optimale ; à l'instant t' le point de phase occupe sur chacune de ces trajectoires une certaine position $x(t')$. En choisissant $u(\nu)$, $t'<\nu<T$, telle qu'elle minimise la fonctionnelle $Q_{t'}=\int_{t'}^{T}G(x(\nu),u(\nu),\nu)\,d\nu$, on définit le prolongement de chacune des trajectoires du faisceau concerné (fig. 14.1). Nous avons déjà dit que les valeurs $\min\limits_{u(\nu)\in\Omega}Q_{t'}=$ $=\min\limits_{u(\nu)\in\Omega}\int_{t'}^{T}G(x(\nu),u(\nu),\nu)\,d\nu$ dépendent de la valeur de la commande u à l'instant t, c'est-à-dire qu'elles seront différentes pour toute trajectoire du faisceau considéré. Voici à $o(\Delta t)$ près les valeurs Q_t^* de la fonctionnelle Q_t sur chacune de ces trajectoires:

$$Q_t^*=Q_t^*(u(t))=G(x(t),u(t),t)\,\Delta t+\min_{u(\nu)\in\Omega}\int_{t'}^{T}G(x(\nu),u(\nu),\nu)\,d\nu.$$

La valeur $u(t)$ qui minimise la fonction $Q_t^*(u(t))$ détermine la trajectoire optimale. Il en résulte que dans la relation (12) la minimisation par rapport à $u(t)$ porte sur toute l'expression entre crochets et non seulement sur son premier terme.

En tenant compte de la notation adoptée

$$S(x',t')=\min_{u(\nu)\in\Omega}\int_{t'}^{T}G(x(\nu),u(\nu),\nu)\,d\nu, \qquad (14.13)$$

récrivons la relation (12)

$$S(x, t) = \min_{u(t)\in\Omega} [G(x(t), u(t), t)\,\Delta t + S(x', t')] + o_1(\Delta t). \quad (14.14)$$

On voit aisément que la relation (14) entre $S(x, t)$ et $S(x', t')$ peut s'obtenir également à partir de (13.30) si l'on tient compte du fait qu'ici aux instants $(N-k)\tau$ et $(N-k+1)\tau$ correspondent les instants t et $t' = t + \Delta t$. Remarquons de plus que la fonction sous le signe somme de la fonctionnelle (13.3) a été désignée par G_1, alors que d'après (13.10), la fonction G de l'expression (13.30) s'écrit $G = \tau G_1$. C'est pourquoi, conformément à (5), il faut remplacer la fonction $G(x(N-k), u(N-k))$ de l'expression (13.30) par la fonction $G(x(t), u(t), t)\,\Delta t = G(x, u, t)\,\Delta t$.

De cette façon, la relation de récurrence (13.30) est remplacée par la relation

$$S(x, t) = \min_{u(t)\in\Omega} [G(x(t), u(t), t)\,\Delta t + S(x', t')] + o_1(\Delta t),$$

où $o_1(\Delta t)$ est un infiniment petit d'ordre supérieur à une grandeur de Δt. Cette grandeur est ajoutée parce que pour déduire la relation (13.30), nous sommes partis de l'équation aux différences (13.7) obtenue à l'aide de la relation approchée (13.6) entachée d'une erreur d'ordre $o_1(\tau)$. L'accroissement de la variable τ de (13.6) est remplacé ici par Δt.

Notons que d'après (1)

$$\begin{aligned} x' = x(t+\Delta t) = x(t) + \frac{dx}{dt}\Delta t + o_2(\Delta t) = \\ = x(t) + f(x(t), u(t), t)\,\Delta t + o_2(\Delta t). \quad (14.15)\end{aligned}$$

Supposons maintenant que la fonction S est continue et qu'elle possède partout des dérivées partielles continues par rapport à toutes ses variables

$$\frac{\partial S}{\partial x_i}\ (i = 1, \ldots, n),\ \frac{\partial S}{\partial t}.$$

Cette hypothèse est très importante du fait qu'elle n'est pas vraie pour de nombreux problèmes pour lesquels les raisonnements qui suivent nécessitent une justification supplémentaire [14].

Dans cette hypothèse, on a en vertu de (15)

$$\begin{aligned} S(x', t') = S(x(t+\Delta t), t+\Delta t) = \\ = S(x(t) + f(x(t), u(t), t)\,\Delta t + o_2(\Delta t), t+\Delta t) = \\ = S(x, t) + \sum_{i=1}^{n} \frac{\partial S(x, t)}{\partial x_i} f_i(x, u, t)\,\Delta t + \frac{\partial S(x, t)}{\partial t}\Delta t + o_3(\Delta t). \quad (14.16)\end{aligned}$$

Si l'on désigne par

$$(\mathrm{grad}\ S, f) = \frac{\partial S}{\partial x_1} f_1 + \frac{\partial S}{\partial x_2} f_2 + \ldots + \frac{\partial S}{\partial x_n} f_n, \quad (14.17)$$

le produit scalaire des vecteurs

$$\operatorname{grad} S = \begin{bmatrix} \frac{\partial S}{\partial x_1} \\ \cdots \\ \frac{\partial S}{\partial x_n} \end{bmatrix}, \quad f(x, u, t) = \begin{bmatrix} f_1(x, u, t) \\ \cdots\cdots \\ f_n(x, u, t) \end{bmatrix},$$

l'expression (16) peut s'écrire

$$S(x', t') = S(x, t) + \langle \operatorname{grad} S(x, t), f(x, u, t)\rangle \Delta t + \\ + \frac{\partial S(x, t)}{\partial t} \Delta t + o_3(\Delta t). \quad (14.18)$$

Portons l'expression (18) dans la relation (14) pour obtenir

$$S(x, t) = \min_{u \in \Omega} [G(x, u, t)\Delta t + S(x, t) + \\ + \langle \operatorname{grad} S(x, t), f(x, u, t)\rangle \Delta t + \frac{\partial S(x, t)}{\partial t} \Delta t + o_3(\Delta t)] + o_1(\Delta t). \quad (14.19)$$

D'après (10) la fonction $S(x, t)$ s'obtient par minimisation de la fonctionnelle $\int_t^T G(x(v), u(v), v)\, dv$ par rapport à u dans l'intervalle $t < v < T$, c'est-à-dire la fonction $S(x, t)$ ne contient déjà plus de u. C'est pourquoi il est possible de faire sortir du signe min les termes $S(x, t)$ et $\frac{\partial S(x, t)}{\partial t} \Delta t$ entre crochets de (19) et mettre l'expression (19) sous la forme

$$S(x, t) = S(x, t) + \frac{\partial S(x, t)}{\partial t} \Delta t + \min_{u \in \Omega} [G(x, u, t)\Delta t + \\ + \langle \operatorname{grad} S(x, t), f(x, u, t)\rangle \Delta t + o_3(\Delta t)] + o_1(\Delta t),$$

ou

$$-\frac{\partial S(x, t)}{\partial t} = \min_{u \in \Omega} [G(x, u, t) + \langle \operatorname{grad} S(x, t), f(x, u, t)\rangle] + \frac{o_4(\Delta t)}{\Delta t}. \quad (14.20)$$

Comme $\lim_{\Delta t \to 0} \frac{o_4(\Delta t)}{\Delta t} = 0$, en passant dans (20) à la limite pour $\Delta t \to 0$, on obtient

$$-\frac{\partial S(x, t)}{\partial t} = \min_{u \in \Omega} [G(x, u, t) + \langle \operatorname{grad} S(x, t), f(x, u, t)\rangle]. \quad (14.21)$$

L'équation (21) [9] s'appelle *équation de Bellman.*

Puisque la minimisation par rapport à u de l'expression entre crochets fait éliminer u du second membre de (21), l'équation (21) sera une équation aux dérivées partielles non linéaire du premier ordre du

type de Jacobi-Hamilton. La fonction u qui minimise l'expression entre crochets de (21) est la commande optimale cherchée u^*. Cette fonction dépend pourtant de grad $S(x, t)$. Ce n'est qu'en résolvant l'équation aux dérivées partielles (21), déjà minimisée par rapport à u, qu'on obtient grad $S(x, t)$ sous une forme explicite, et par suite, la forme explicite de la commande optimale u^*.

Voici les *conditions aux limites* de l'équation (21) : pour la fonctionnelle (5), $S(x(T), T) = 0$; pour la fonctionnelle (13.3), $S(x(T), T) = \varphi(x(T))$.

Soulignons maintenant que pour déduire l'équation (21) le point de départ était la trajectoire optimale déjà trouvée du système (1). Dans l'hypothèse de la fonction $S(x, t)$ lisse, l'équation (21) a lieu pour toute trajectoire optimale. Ainsi, sous cette hypothèse, l'équation de Bellman donne les conditions *nécessaires* d'optimalité. La question de la suffisance de ces conditions demande une étude spéciale.

E x e m p l e 1. Considérons le système décrit par les équations [85] :

$$\frac{dx_1}{dt} = ux_1 + x_2,$$
$$\frac{dx_2}{dt} = u_2.$$

Supposons que la fonction sous le signe somme de la fonctionnelle (5) s'écrit

$$G = G(x_1, x_2, t).$$

Dans notre exemple, (1) entraîne

$$f_1(x, u) = ux_1 + x_2,$$
$$f_2(x, u) = u^2.$$

Ici l'équation de Bellman (21) est de la forme

$$-\frac{\partial S}{\partial t} = \min_u K(x_1, x_2, u),$$

où

$$K(x_1, x_2, u) = G(x_1, x_2, t) + \frac{\partial S}{\partial x_1}(ux_1 + x_2) + \frac{\partial S}{\partial x_2}u^2.$$

Comme

$$\frac{dK}{du} = \frac{\partial S}{\partial x_1}x_1 + 2\frac{\partial S}{\partial x_2}u, \quad \frac{d^2K}{du^2} = 2\frac{\partial S}{\partial x_2},$$

dans le cas où $\frac{\partial S}{\partial x_2} > 0$, la fonction K aura son minimum pour $u = u^*$, avec

$$u^* = -\frac{1}{2}\frac{\partial S}{\partial x_1}x_1\frac{1}{\frac{\partial S}{\partial x_2}}.$$

La valeur minimale de $[K(x_1, x_2, u)]_{u=u^*}$ devient

$$\min_u K(x_1, x_2, u) = K(x_1, x_2, u^*) = G(x_1, x_2, t) + \frac{\partial S}{\partial x_1} x_2 - \frac{1}{4} x_1^2 \left(\frac{\partial S}{\partial x_1}\right)^2 \frac{1}{\frac{\partial S}{\partial x_2}}.$$

Ainsi, dans l'exemple considéré, l'équation de Bellman (21) se met sous la forme

$$-\frac{\partial S}{\partial t} = G(x_1, x_2, t) + \frac{\partial S}{\partial x_1} x_2 - \frac{1}{4} x_1^2 \left(\frac{\partial S}{\partial x_1}\right)^2 \frac{1}{\frac{\partial S}{\partial x_2}}.$$

Exemple 2. Dans l'exemple précédent la commande u n'était pas limitée par des contraintes. Considérons maintenant le système décrit par les équations

$$\frac{dx_1}{dt} = u_1 x_1 + x_2,$$

$$\frac{dx_2}{dt} = u_2.$$

Supposons que la fonction sous le signe somme de la fonctionnelle (5) est

$$G = G(x_1, x_2, t).$$

Imposons aux commandes u_1, u_2 les contraintes

$$-1 \leqslant u_i \leqslant 1 \qquad (i = 1, 2).$$

D'après (1) on a

$$f_1(x, u) = u_1 x_1 + x_2,$$
$$f_2(x, u) = u_2.$$

L'équation de Bellman (21) devient

$$-\frac{\partial S}{\partial t} = \min_{\substack{-1 \leqslant u_1 \leqslant 1 \\ -1 \leqslant u_2 \leqslant 1}} K(x_1, x_2, u),$$

où

$$K(x_1, x_2, u) = G(x_1, x_2, t) + \frac{\partial S}{\partial x_1}(u_1 x_1 + x_2) + \frac{\partial S}{\partial x_2} u_2.$$

La commande optimale u_i^* qui minimise la fonction $K(x_1, x_2, u)$ s'écrit

$$u_1^* = -\operatorname{sign}\left(x_1 \frac{\partial S}{\partial x_1}\right), \quad u^* = -\operatorname{sign}\frac{\partial S}{\partial x_2}.$$

Ainsi, sous les contraintes $|u_i| \leqslant 1$ $(i=1, 2)$ imposées à la commande, la valeur minimale de $[K(x_1, x_2, u)]_{u=u^*}$ est de la forme

$$[K(x_1, x_2, u)]_{u=u^*} = G(x_1, x_2, t) - \left|x_1 \frac{\partial S}{\partial x_1}\right| + x_2 \frac{\partial S}{\partial x_1} - \left|\frac{\partial S}{\partial x_2}\right|.$$

L'équation de Bellman (21) est ici

$$-\frac{\partial S}{\partial t} = G(x_1, x_2, t) - \left|x_1 \frac{\partial S}{\partial x_1}\right| + x_2 \frac{\partial S}{\partial x_1} - \left|\frac{\partial S}{\partial x_2}\right|.$$

Remarquons que sous les contraintes $-1 \leqslant u_i \leqslant 1$ ($i = 1, 2$) imposées à la commande, la fonction $S(x_1, x_2, t)$ n'est pas partout lisse (cf. ci-dessous § 15). L'équation de Bellman obtenue dans ce dernier exemple ne doit donc être envisagée que dans le domaine où la fonction S est continue et possède des dérivées partielles continues par rapport à toutes ses variables.

2. Problème à extrémité de trajectoire fixe et à temps libre. Amener le système de commande régi par les équations différentielles

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r) \qquad (j = 1, \ldots, n) \tag{14.22}$$

auxquelles est équivalente l'équation vectorielle

$$\frac{dx}{dt} = f(x, u), \tag{14.23}$$

où x, u et f sont des vecteurs de la forme

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u) = \begin{bmatrix} f_1(x, u) \\ \cdot \quad \cdot \quad \cdot \\ f_n(x, u) \end{bmatrix},$$

du point $x(t_0) = x^0$ de l'espace de phase X au point donné x^*. L'instant t_1 lorsque le point représentatif tombe en x^* n'est pas fixé à l'avance.

La commande $u = u(t)$ doit vérifier la condition

$$u \in \Omega, \tag{14.24}$$

et elle doit être choisie de façon que la fonctionnelle

$$Q = \int_{t_0}^{t_1} G(x(\nu), u(\nu))\, d\nu \tag{14.25}$$

soit minimisée.

Admettons que la commande qui satisfait à ces conditions, la trajectoire qui lui correspond et l'intervalle de temps $t_1 - t_0$ sont *optimaux*.

La plus petite valeur possible de Φ de la fonctionnelle Q obtenue sous ces conditions est une fonction de l'état initial x^0 du système

$$\Phi = \Phi(x^0) = \Phi(x_1^0, x_2^0, \ldots, x_n^0). \tag{14.26}$$

Sous l'hypothèse que la fonction Φ est continue et possède partout des dérivées partielles continues par rapport à toutes ses variables, on peut obtenir une équation aux dérivées partielles (équation de Bellman) vérifiée par cette fonction.

Par définition

$$\Phi(x^0) = \Phi(x(t_0)) = \min_{u(\nu) \in \Omega} \int_{t_0}^{t_1} G(x(\nu), u(\nu))\, d\nu. \tag{14.27}$$

L'intégrale du second membre de (27) peut s'écrire

$$\int_{t_0}^{t_1} G(x(\nu),\ u(\nu))\,d\nu = \int_{t_0}^{t_0+\Delta t} G(x(\nu),\ u(\nu))\,d\nu + \\ + \int_{t_0+\Delta t}^{t_1} G(x(\nu),\ u(\nu))\,d\nu. \quad (14.28)$$

Pour de petites valeurs de Δt, l'expression (28) peut être mise sous la forme

$$\int_{t_0}^{t_1} G(x(\nu),\ u(\nu))\,d\nu = G(x(t_0),\ u(t_0))\,\Delta t + \\ + o_1(\Delta t) + \int_{t_0+\Delta t}^{t_1} G(x(\nu),\ u(\nu))\,d\nu, \quad (14.29)$$

où l'on suppose que la fonction $u(t)$ est continue dans l'intervalle semi-ouvert $[t_0,\ t_0+\Delta t)$. L'expression (27) devient

$$\Phi(x^0) = \min_{u(t_0)\in\Omega} \Big[G(x(t_0),\ u(t_0))\,\Delta t + \\ + \min_{u(\nu)\in\Omega} \int_{t_0+\Delta t}^{t_1} G(x(\nu),\ u(\nu))\,d\nu \Big] + o_2(\Delta t). \quad (14.30)$$

La présence du signe min devant les crochets de (30) est justifiée par les mêmes raisonnements que ceux qui ont été donnés pour déduire la relation (12).

Le deuxième terme entre crochets de (30) est $\Phi(x(t_0+\Delta t))$ et cette expression peut s'écrire

$$\Phi(x(t_0)) = \min_{u(t_0)\in\Omega} [G(x(t_0),\ u(t_0))\,\Delta t + \Phi(x(t_0+\Delta t))] + o_2(\Delta t). \quad (14.31)$$

Puisque d'après (23)

$$x(t_0+\Delta t) = x(t_0) + \left(\frac{dx}{dt}\right)_{t=t_0} \Delta t + o_3(\Delta t) = \\ = x(t_0) + f(x(t_0),\ u(t_0))\,\Delta t + o_3(\Delta t), \quad (14.32)$$

l'hypothèse suivant laquelle la fonction $\Phi(x^0)$ est lisse, conduit à

$$\Phi(x(t_0+\Delta t)) = \Phi(x(t_0)) + \\ + \sum_{i=1}^{n} \left[\frac{\partial\Phi(x)}{\partial x_i}\right]_{x=x(t_0)} f_i(x(t_0),\ u(t_0))\,\Delta t + o_4(\Delta t). \quad (14.33)$$

Portons maintenant l'expression (33) dans (31). Si l'on tient compte du fait que la fonction $\Phi(x(t_0))$, résultant de la minimisation de la fonctionnelle (25), ne contient plus de u, $\Phi(x(t_0))$ de (31) peut être sortie du signe min. Il vient alors

$$\min_{u(t_0)\in\Omega}\Big\{G(x(t_0),\, u(t_0))\,\Delta t + \sum_{i=1}^{n}\Big[\frac{\partial\Phi(x)}{\partial x_i}\Big]_{x=x(t_0)} f_i(x(t_0,\, u(t_0))\ \Delta t\Big\} + o(\Delta t) = 0. \quad (14.34)$$

On peut adopter comme état initial tout état courant $x(t)$ et récrire respectivement la relation (34)

$$\min_{u\in\Omega}\Big[G(x,\, u) + \sum_{i=1}^{n}\frac{\partial\Phi(x)}{\partial x_i}f_i(x,\, u)\Big] + \frac{o(\Delta t)}{\Delta t} = 0. \quad (14.35)$$

Puisque $\lim\limits_{\Delta t\to 0}\frac{o(\Delta t)}{\Delta t} = 0$, en passant dans (35) à la limite pour $\Delta t \to 0$, on obtient

$$\min_{u\in\Omega}\Big[G(x,\, u) + \sum_{i=1}^{n}\frac{\partial\Phi(x)}{\partial x_i}f_i(x,\, u)\Big] = 0. \quad (14.36)$$

L'équation (36) est précisément l'*équation de Bellman* relative au problème étudié à extrémité de trajectoire fixe et à temps libre.

La solution de l'équation (36) doit satisfaire à la condition aux limites

$$\Phi(x_1^*,\ x_2^*,\ \ldots,\ x_n^*) = 0.$$

3. Problème de la réponse en temps minimal.[1] Parmi les problèmes dans lesquels l'instant t_1 où le processus de commande prend fin n'est pas fixé à l'avance, il y a également celui de la *réponse en temps minimal*.

Considérons le système décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = f(x,\, u), \quad (14.37)$$

où x, u et f sont des vecteurs de la forme

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x,\, u) = \begin{bmatrix} f_1(x,\, u) \\ \cdots\cdot \\ f_n(x,\, u) \end{bmatrix}.$$

Les commandes u_i sont soumises aux contraintes

$$u \in \Omega, \quad (14.38)$$

qui, en particulier, peuvent s'écrire

$$-m_i \leqslant u_i \leqslant m_i \qquad (i = 1, \ldots, r). \tag{14.39}$$

Chercher la commande optimale $u \in \Omega$ qui *en temps minimal* T recale le système de l'état initial $x(t_0) = x^0$ à l'état $x(t_1) = x(t_0 + T) = 0$, c'est-à-dire à l'origine des coordonnées de l'espace de phase.

Le temps minimal T nécessaire pour que la commande $u \in \Omega$ ramè ne le système du point $x(t_0) = x^0$ au point $x(t_0 + T) = 0$, est fonction de l'état initial du système

$$T = T(x^0) = T(x_1^0, x_2^0, \ldots, x_n^0), \tag{14.40}$$

et le problème de la réponse en temps minimal est un cas particulier du problème, étudié au point 2, qui consiste à minimiser la fonctionnelle (25)

$$Q = \int_{t_0}^{t_1} G(x(\nu), u(\nu))\, d\nu.$$

En effet, en posant

$$G(x(t), u(t)) \equiv 1, \tag{14.41}$$

on trouve que la fonctionnelle Q est le temps nécessaire pour recaler le système de l'état initial $x(t_0) = x^0$ à l'état $x(t_1) = 0$:

$$Q = t_1 - t_0; \tag{14.42}$$

par conséquent,

$$T = \min_{u \in \Omega} Q = \min_{u \in \Omega} (t_1 - t_0). \tag{14.43}$$

Si l'on tient compte du fait que pour l'état initial on peut prendre tout état courant $x(t)$ et si l'on suppose que la fonction $T(x_1, x_2, \ldots, x_n)$ soit continue et possède partout des dérivées partielles continues par rapport à toutes ses variables, on obtient d'après (36) et (41) l'équation différentielle aux dérivées partielles du premier ordre vérifiée par la fonction $T(x_1, \ldots, x_n)$:

$$\min_{u \in \Omega} \left[\sum_{i=1}^{n} \frac{\partial T}{\partial x_i} f_i(x, u) \right] = -1$$

ou

$$\min_{u \in \Omega} [\langle \operatorname{grad} T, f(x, u) \rangle] = -1. \tag{14.44}$$

L'équation (44) constitue précisément l'*équation de Bellman* du problème de la *réponse en temps minimal*.

La minimisation par rapport à u de l'expression entre crochets du premier membre de (44) sous la condition (38) permet de définir

la commande optimale u^* qui figure sous la forme de la fonction de $\frac{\partial T}{\partial x_i}$ $(i = 1, \ldots, n)$. En portant dans (44) cette valeur u^* on obtient une équation aux dérivées partielles ne contenant pas de u. La solution de cette équation doit satisfaire à la condition aux limites

$$T(0, 0, \ldots, 0) = 0. \tag{14.45}$$

Si l'on réussit à la résoudre, on détermine par là la commande optimale $u^* = u^*(x)$ sous la forme d'une fonction explicite de coordonnées de phase du système.

Malheureusement jusqu'à présent l'équation (44) n'a été résolue sous une forme fermée que pour les cas les plus simples.

Exemple. Considérons à titre d'exemple le problème de la réponse en temps minimal d'un système linéaire décrit par les équations différentielles scalaires

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} A_{jk}x_k + \sum_{l=1}^{r} B_{jl}u_l \qquad (j = 1, \ldots, n), \tag{14.46}$$

où les commandes u_l sont soumises aux contraintes

$$-m_l \leqslant u_l \leqslant m_l \qquad (l = 1, \ldots, r). \tag{14.47}$$

Au système (46) est équivalente l'équation différentielle vectorielle

$$\frac{dx}{dt} = Ax + Bu, \tag{14.48}$$

où

$$x = \begin{bmatrix} x_1 \\ \ldots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} \ldots A_{1n} \\ \ldots\ldots \\ A_{n1} \ldots A_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} \ldots B_{1r} \\ \ldots\ldots \\ B_{n1} \ldots B_{nr} \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \ldots \\ u_r \end{bmatrix}. \tag{14.49}$$

Conformément à (37) dans l'exemple considéré

$$f_j(x, u) = \sum_{k=1}^{n} A_{jk}x_k + \sum_{l=1}^{r} B_{jl}u_l \qquad (j = 1, \ldots, n), \tag{14.50}$$

et l'équation de Bellman devient

$$\min_{u \in \Omega} \left[\sum_{j=1}^{n} \frac{\partial T}{\partial x_j} \left(\sum_{k=1}^{n} A_{jk}x_k + \sum_{l=1}^{r} B_{jl}u_l \right) \right] = -1. \tag{14.51}$$

L'équation (51) peut s'écrire

$$\min_{u \in \Omega} \left[\sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial T}{\partial x_j} A_{jk}x_k + \sum_{l=1}^{r} \left(\sum_{j=1}^{n} \frac{\partial T}{\partial x_j} B_{jl} \right) u_l \right] = -1. \tag{14.52}$$

Sous les contraintes (47), l'expression entre crochets du premier membre de (52) est minimisée par les commandes

$$u_l = -m_l \operatorname{sign} \sum_{j=1}^{n} \frac{\partial T}{\partial x_j} B_{jl} \qquad (l=1, \ldots, r). \tag{14.53}$$

Pour les commandes (53) l'équation de Bellman (52) se met sous la forme

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial T}{\partial x_j} A_{jk} x_k - \sum_{l=1}^{r} \left| \sum_{j=1}^{n} \frac{\partial T}{\partial x_j} B_{jl} \right| m_l = -1. \tag{14.54}$$

Nous avons obtenu l'équation différentielle aux dérivées partielles non linéaire du premier ordre (54) ne contenant pas de u_l dont la solution doit satisfaire aux conditions aux limites

$$T(0, 0, \ldots, 0) = 0. \tag{14.55}$$

L'équation de Bellman (54) que nous venons d'obtenir doit être étudiée dans le domaine où la fonction T est continue et où elle possède des dérivées partielles continues par rapport à toutes ses variables.

§ 15. Conditions suffisantes d'optimalité et justification de la méthode de programmation dynamique pour les systèmes continus

La méthode de programmation dynamique est bien justifiée pour des systèmes régis par les équations aux différences finies.

Pour des systèmes décrits par les équations différentielles ordinaires (c'est-à-dire pour des systèmes continus), la justification correcte de la méthode de programmation dynamique nécessiterait la démonstration de l'existence de la fonction lisse $\Phi = \Phi(x_1, \ldots, x_n)$ définie par une équation non linéaire aux dérivées partielles (14.36) qui est l'équation de Bellman (il en est de même de la fonction $T(x)$ définie par (14.44)), du fait que pour déduire cette équation nous avons supposé que la fonction Φ est continue et possède partout des dérivées partielles continues par rapport à toutes ses variables. Cette restriction ne résulte pas de la position du problème et constitue une contrainte qui, toutefois, n'est pas respectée dans de nombreux problèmes, même simples, dans lesquels [15] la fonction $\Phi(x_1, \ldots, x_n)$ n'est pas lisse.

Le problème de justification de la méthode de programmation dynamique pour des systèmes à action continue a été résolu par V. Boltianski dans ses travaux [14] à l'exposé desquels nous allons passer.

1. Position du problème. Interprétation géométrique de l'équation de Bellman dans le problème de la réponse en temps minimal. Considérons le système commandé décrit par les équations différentielles scalaires

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r) \qquad (j=1, \ldots, n). \tag{15.1}$$

Ce système peut s'écrire sous une forme vectorielle

$$\frac{dx}{dt} = f(x, u), \tag{15.2}$$

où x, u et f sont les vecteurs

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u) = \begin{bmatrix} f_1(x, u) \\ \cdots\cdot \\ f_n(x, u) \end{bmatrix}. \tag{15.3}$$

La commande u est soumise aux contraintes

$$u \in \Omega. \tag{15.4}$$

Considérons le problème de la réponse en temps minimal qui consiste à chercher la commande optimale $u \in \Omega$, c'est-à-dire celle qui ramène le système de l'état initial $x(t)$ à l'état $x(t+T) = 0$ en un temps minimal T.

L'état final du système est $x(t+T)$, c'est-à-dire le point de l'espace de phase en lequel le système doit être amené est fixé: $x(t+T) = 0$. Comme état initial nous adoptons un état quelconque, c'est-à-dire n'importe quel point x de l'espace de phase X.

Le temps minimal T nécessaire à la commande u vérifiant la condition $u \in \Omega$ pour amener le système du point $x(t)$ au point $x(t+T) = 0$ est fonction de l'état initial x:

$$T = T(x) = T(x_1, \ldots, x_n).$$

Nous avons montré au § 14 en supposant que la fonction $T(x_1, \ldots$ $\ldots, x_n)$ est continue et possède partout (sauf au point final $x = 0$) des dérivées partielles continues $\frac{\partial T}{\partial x_1}, \frac{\partial T}{\partial x_2}, \ldots, \frac{\partial T}{\partial x_n}$, que cette fonction vérifie l'équation aux dérivées partielles (14.44)

$$\min_{u \in \Omega} \left[\sum_{i=1}^{n} \frac{\partial T(x)}{\partial x_i} f_i(x, u) \right] = -1.$$

La commande qui minimise l'expression entre crochets est *optimale*, nous l'avons désignée par u^*.

Dans ce qui suit au lieu de $T(x)$ il est plus commode d'utiliser la fonction $\omega(x)$ qui ne diffère de $T(x)$ que par le signe

$$\omega(x) = -T(x). \tag{15.5}$$

La relation (14.44) entraîne

$$\max_{u \in \Omega} \left[-\sum_{i=1}^{n} \frac{\partial T(x)}{x_i} f_i(x, u) \right] = 1. \tag{15.6}$$

Compte tenu de (5), la relation (6) peut s'écrire

$$\max_{u\in\Omega}\left[\sum_{i=1}^{n}\frac{\partial\omega(x)}{\partial x_i}f_i(x, u)\right]=1. \tag{15.7}$$

La fonction $\omega(x)$ est définie dans l'espace de phase X tout entier. En vertu des hypothèses qui ont permis d'obtenir la relation (14.44), la relation (7) a lieu sous l'hypothèse que la fonction $\omega(x)$ est continue et possède partout (sauf au point final $x = 0$) des dérivées partielles continues $\frac{\partial\omega}{\partial x_1}, \frac{\partial\omega}{\partial x_2}, \ldots, \frac{\partial\omega}{\partial x_n}$. La relation (7) est une équation de Bellman pour le problème de la réponse en temps minimal.

Pour *toute* commande de l'ensemble des commandes limitées par la condition $u\in\Omega$ (y compris la commande optimale) qui recalent le système du point x au point $x = 0$, on a, conformément à (7), la relation

$$\sum_{i=1}^{n}\frac{\partial\omega(x)}{\partial x_i}f_i(x, u)\leqslant 1 \qquad (u\in\Omega). \tag{15.8}$$

Comme $x = x(t)$, $u = u(t)$, la relation

$$\sum_{i=1}^{n}\frac{\partial\omega(x(t))}{\partial x_i}f_i(x(t), u(t))=1 \tag{15.9}$$

est observée pendant toute la durée du mouvement pour chaque trajectoire optimale, c'est-à-dire pendant toute la durée de passage du point initial x au point final $x = 0$.

Voici l'interprétation géométrique qu'on peut donner à l'équation de Bellman dans le problème de la réponse en temps minimal. Puisque d'après (1)

$$f_i(x, u)=\frac{dx_i}{dt}, \tag{15.10}$$

l'équation de Bellman (14.44) peut s'écrire

$$\min_{u\in\Omega}\left[\sum_{i=1}^{n}\frac{\partial T(x)}{\partial x_i}\frac{dx_i}{dt}\right]=\min_{u\in\Omega}\frac{dT(x(t))}{dt}=-1, \tag{15.11}$$

c'est-à-dire lorsqu'on se déplace suivant la trajectoire du mouvement optimale, l'intervalle de temps qui reste pour atteindre l'origine des coordonnées diminue de façon que $-\frac{dT}{dt}=1$.

D'après (11)

$$\min_{u\in\Omega}\left[\left\langle \operatorname{grad} T, \frac{dx}{dt}\right\rangle\right]=-1. \tag{15.12}$$

L'expression entre crochets est le produit scalaire du vecteur gradient de la fonction $T(x)$ par le vecteur vitesse de phase du point représentatif.

Imaginons une famille de surfaces $T(x) = \text{const}$. On peut les appeler surfaces isochrones du fait que le temps minimal nécessaire pour atteindre l'origine à partir d'un point quelconque de la surface $T(x) = C$ est égal à la constante C. Au point x, le vecteur grad T est dirigé suivant la normale à la surface $T(x) = \text{const}$ qui passe par ce point. La relation (12) montre que pour la commande optimale $u = u^*$ la projection de la vitesse de phase $\frac{dx}{dt}$ sur la direction négative de la normale à la surface $T(x) = \text{const}$, qui passe par le point x, est à tout instant positive, c'est-à-dire le point représentatif se déplace toujours dans le sens des valeurs décroissantes de $T(x)$.

L'hypothèse sur l'existence des dérivées partielles continues $\frac{\partial T(x)}{\partial x_i}$ $(i = 1, \ldots, n)$ en tout point x équivaut à l'hypothèse sur les surfaces lisses $T(x) = \text{const}$, c'est-à-dire sur l'unicité de la normale en tout point de ces surfaces. Les points où les dérivées partielles $\frac{\partial T(x)}{\partial x_i}$ $(i = 1, \ldots, n)$ n'existent pas sont des points singuliers des surfaces $T(x) = \text{const}$, et en ces points la direction de la normale à la surface $T(x) = \text{const}$ n'est pas définie.

2. Condition suffisante d'optimalité du problème de la réponse en temps minimal. Passons à la recherche des conditions suffisantes d'optimalité. La méthode proposée par V. Boltianski consiste à étudier la fonction $\omega(x)$ d'un autre point de vue et en particulier, à choisir comme point de départ la relation (8). Cette relation permet d'apprécier le temps nécessaire pour qu'un système passe d'un état initial quelconque en un état final donné. Admettons de plus que les seconds membres des équations différentielles (1) sont définies, continues et possèdent des dérivées continues $\frac{\partial f_i}{\partial x_j}$ en tout point $x \in X$ quand $u \in \Omega$.

L e m m e 1. *Supposons donnée sur un ensemble ouvert D de l'espace X une fonction continue et continûment dérivable $\omega(x) = \omega(x_1, x_2, \ldots, x_n)$ telle que pour tout $x \in D$, $u \in \Omega$, elle vérifie l'inégalité de la forme* (8)

$$\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \leqslant 1. \tag{15.13}$$

Alors, si $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ est une commande admissible (c'est-à-dire si $u(t) \in \Omega$) qui ramène le point représentatif de la position x^0 en position x^*, la trajectoire correspondante $x(t)$ reposant tout entière dans l'ensemble D, le temps $t_1 - t_0$ de passage de x^0 en x^* est évalué

par l'inégalité

$$t_1 - t_0 \geqslant \omega(x^*) - \omega(x^0). \tag{15.14}$$

D é m o n s t r a t i o n. Soient $\tau_1, \tau_2, \ldots, \tau_k$ les points de discontinuité des commandes $u(t)$; en outre

$$t_0 < \tau_1 < \tau_2 < \ldots < \tau_k < t_1.$$

Introduisons encore les notations

$$t_0 = \tau_0, \qquad t_1 = \tau_{k+1}.$$

Les points de discontinuité des commandes ont en général lieu si, par exemple, les contraintes (4) imposées à la commande sont de la forme

$$-m_i \leqslant u_i \leqslant m_i \qquad (i = 1, \ldots, r).$$

Dans chaque intervalle (τ_0, τ_1), (τ_1, τ_2), $\ldots$, (τ_k, τ_{k+1}) la fonction $u(t)$ est continue.

D'après (8) et (1),

$$\sum_{i=1}^{n} \frac{\partial \omega(x(t))}{\partial x_i} f_i(x(t), u(t)) = \sum_{i=1}^{n} \frac{\partial \omega(x(t))}{\partial x_i} \frac{\partial x_i(t)}{dt} = \frac{d\omega(x(t))}{dt} \leqslant 1. \tag{15.15}$$

A l'intérieur de chaque intervalle (τ_i, τ_{i+1}) $(i = 0, 1, \ldots, k)$ la fonction $\omega(x(t))$ est continue et possède, en vertu de la condition du lemme, une dérivée continue par rapport à t, qui vérifie l'inégalité (15)

$$\frac{d\omega(x(t))}{dt} \leqslant 1.$$

Il en résulte que

$$\omega(x(\tau_{i+1})) - \omega(x(\tau_i)) \leqslant \tau_{i+1} - \tau_i \qquad (i = 0, 1, \ldots, k).$$

En additionnant les premiers et les seconds membres de ces relations écrites pour les intervalles (τ_i, τ_{i+1}) $(i = 0, 1, \ldots, k)$, on a

$$\omega(x(\tau_{k+1})) - \omega(x(\tau_0)) \leqslant \tau_{k+1} - \tau_0,$$

ce qui coïncide avec la relation (14). Ainsi le lemme est démontré.

L'adjonction d'une constante à la fonction $\omega(x)$ ne modifiant pas la relation (8), on peut poser que la fonction $\omega(x)$ observe la condition

$$\omega(x^*) = 0, \tag{15.16}$$

ce qui correspond à la condition aux limites (14.45).

Dans ces conditions l'inégalité (14) devient

$$t_1 - t_0 \geqslant -\omega(x^0). \tag{15.17}$$

Notons que l'équation (7), qui est une équation de Bellman du problème de la réponse en temps minimal, a été obtenue précédemment comme la condition *nécessaire* d'optimalité.

Nous avons démontré ci-dessus sous les mêmes hypothèses de continuité et de dérivabilité continue de la fonction $\omega(x)$ et sous la condition $\omega(x^*) = 0$, que si la relation (8) a lieu, la commande qui ramène le système de x^0 en x^* en un temps $-\omega(x^0)$ est optimale. Il s'ensuit de (17) que cette transition ne peut être réalisée plus vite. Nous avons démontré donc que la relation (7) est une condition *suffisante* d'optimalité.

Ainsi dans le cas où $\omega(x)$ est une fonction continue et continûment dérivable, l'équation de Bellman (7) est la condition *nécessaire et suffisante* d'optimalité.

Nous avons déjà dit que dans de nombreux problèmes la restriction imposant la fonction $\omega(x)$ continûment dérivable n'était pas respectée. Considérons à

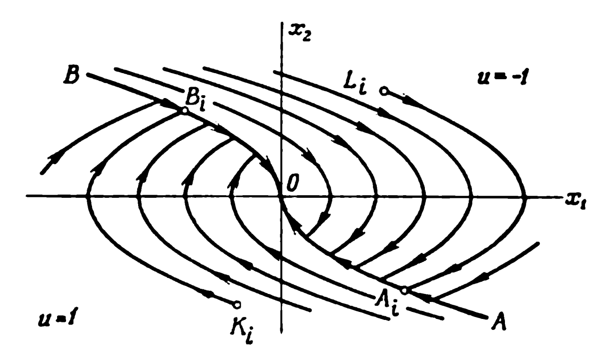


Fig. 15.1

titre d'exemple d'un tel problème celui de la construction des commandes à temps optimal dans le cas d'un système régi par les équations

$$\frac{dx_1}{dt} = x_2,$$

$$\frac{dx_2}{dt} = u,$$

où la commande admissible est limitée par la contrainte

$$-1 \leqslant u \leqslant 1.$$

On peut montrer (cf. p. 273) que l'arc AO (fig. 15.1) de la parabole $x_1 = \frac{1}{2} x_2^2$ et l'arc OB de la parabole $x_1 = -\frac{1}{2} x_2^2$ sont les lignes de commutation de la commande optimale. Au-dessus de la ligne AOB la commande doit être $u = -1$. Le point représentatif suit l'arc L_iA_i de la parabole $x_1 = -\frac{1}{2} x_2^2 + c_i$. Le point

A_i de l'intersection de cette parabole avec la ligne AO est celui de la commutation de la commande qui devient $u = 1$. Le mouvement ultérieur s'effectue suivant la ligne A_iO jusqu'à l'origine des coordonnées.

Au-dessous de la ligne AOB il faut poser $u = 1$. Le point représentatif se déplace suivant l'arc K_iB_i de la parabole $x_1 = \frac{1}{2} x_2^2 + c_i$. Le point B_i de l'intersection de cette parabole avec la courbe BO est celui de la commutation de la commande qui devient $u = -1$. Le mouvement se poursuit suivant la ligne B_iO jusqu'à l'origine des coordonnées.

La trajectoire L_iA_iO et la trajectoire K_iB_iO sont des trajectoires parcourues en un temps minimal du point représentatif dont les positions initiales reposent sur les lignes L_iA_i et K_iB_i respectivement.

Nous montrons ci-dessous (p. 277) que la fonction $\omega(x) = -T(x)$ est continue dans le plan de phase x_1x_2 tout entier. Les dérivées partielles $\frac{\partial\omega(x)}{\partial x_1}$ et $\frac{\partial\omega(x)}{\partial x_2}$ existent en tout point du plan de phase x_1x_2, sauf les points de la ligne de commutation AOB. Les dérivées partielles $\frac{\partial\omega(x)}{\partial x_1}$ et $\frac{\partial\omega(x)}{\partial x_2}$ n'existent en aucun point de la ligne de commutation AOB.

Supposons dans le cas général que nous avons découpé dans l'ensemble D sur lequel est donnée la fonction $\omega(x)$ un ensemble M (« ensemble singulier » de la fonction $\omega(x)$), que la fonction $\omega(x)$ est continue sur l'ensemble D tout entier et qu'elle ne possède des dérivées partielles continues $\frac{\partial\omega(x)}{\partial x_i}$ $(i = 1, \ldots, n)$ qu'aux points de l'ensemble D qui n'appartiennent pas à M.

Lemme 2. *Soit D un ensemble ouvert de l'espace de phase X et M un certain ensemble contenu dans D. Supposons donnée sur l'ensemble D une fonction continue $\omega(x) = \omega(x_1, x_2, \ldots, x_n)$ possédant en dehors de M des dérivées partielles continues et vérifiant l'inégalité* (8)

$$\sum_{i=1}^{n} \frac{\partial\omega(x)}{\partial x_i} f_i(x, u) \leqslant 1.$$

Soit, ensuite, $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ la commande admissible qui ramène le point représentatif de la position x^0 en position x^* ; de plus, la trajectoire correspondante $x(t)$ se trouve tout entière dans D et ne coupe l'ensemble M qu'en un nombre fini d'instants. Alors, le critère (14)

$$t_1 - t_0 \geqslant \omega(x^*) - \omega(x^0)$$

est encore vrai.

Démonstration.

1°. Remarquons que le critère (14) est encore vrai si la trajectoire $x(t)$ ne coupe l'ensemble M qu'aux instants t_0, t_1 ou à l'un de ces instants. En effet, c'est le cas qui correspond complètement à la démonstration du lemme 1 du fait que la fonction $\omega(x)$ est encore continue, et lorsque t se trouve à l'intérieur des intervalles (τ_i, τ_{i+1})

(dont les limites sont des instants de discontinuité des commandes $u(t)$), le point $\omega(x(t))$ se situe en dehors de M et donc dans ces intervalles les dérivées $\frac{\partial\omega(x(t))}{\partial x_i}$ existent bien et sont continues.

2°. Soient maintenant $\theta_1, \theta_2, \ldots, \theta_s$ tous les instants de l'intervalle (t_0, t_1), où la trajectoire $x(t)$ coupe l'ensemble M; de plus, $t_0 < \theta_1 < \theta_2 < \ldots < \theta_s < t_1$. Introduisons encore les notations

$$t_0 = \theta_0, \qquad t_1 = \theta_{s+1}.$$

A chacun des intervalles $[\theta_i, \theta_{i+1}]$ $(i = 0, 1, \ldots, s)$ on peut appliquer la remarque 1° et obtenir

$$\theta_{i+1} - \theta_i \geqslant \omega(x(\theta_{i+1})) - \omega(x(\theta_i)) \qquad (i = 0, 1, \ldots, s).$$

En additionnant les premiers et les seconds membres de ces relations associés aux intervalles de temps $[\theta_i, \theta_{i+1}]$ $(i = 0, 1, \ldots, s)$, on a la relation

$$\theta_{s+1} - \theta_0 \geqslant \omega(x(\theta_{s+1})) - \omega(x(\theta_0)),$$

qui coïncide avec l'inégalité (14) imposée par la condition du lemme. Le lemme est ainsi démontré.

Notons que la situation qui a lieu dans le problème de la réponse en temps minimal (cf. l'exemple de la fig. 15.1) n'obéit pas aux conditions du lemme 2.

En effet, dans cet exemple (fig. 15.1, 15.2) ce ne sont pas des points isolés mais un arc de la trajectoire optimale qui se situe sur la ligne de commutation AOB, c'est-à-dire, dans les termes du lemme 2, tout un arc de la trajectoire optimale se situe dans l'ensemble M. La condition du lemme 2 n'admet l'intersection de la trajectoire optimale avec l'ensemble M qu'en un nombre fini d'instants. Or, dans le cas considéré, la trajectoire optimale se trouve dans l'ensemble M pendant un certain intervalle de temps.

Cette difficulté est levée par le lemme 3 qui suit.

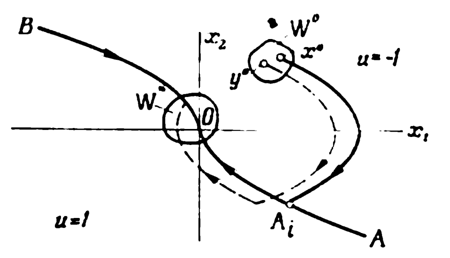


Fig. 15.2

L e m m e 3. *Conservons les hypothèses du lemme 2 sur la fonction* ω. *Soit* $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ *la commande admissible qui ramène le point représentatif de la position* x^0 *en position* x^*; *en outre, la trajectoire correspondante* $x(t)$ *est située tout entière dans* D. *Supposons encore qu'il existe aussi près que l'on veut de* x^0 *un point* y^0 *tel que la trajectoire* $y(t)$ *issue de* y^0 *et associée à la même commande* $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ *ne coupe* M *qu'en un nombre fini d'instants. Alors le critère* (14)

$$t_1 - t_0 \geqslant \omega(x^*) - \omega(x^0)$$

est encore vrai.

D é m o n s t r a t i o n. Choisissons un nombre arbitraire $\varepsilon > 0$ et soient W^0 et W^* (fig. 15.2) respectivement les voisinages des points x^0 et x^* tels que

$$\left.\begin{array}{l} |\,\omega(x) - \omega(x^0)\,| < \varepsilon \quad \text{pour} \quad x \in W^0, \\ |\,\omega(x) - \omega(x^*)\,| < \varepsilon \quad \text{pour} \quad x \in W^*. \end{array}\right\} \qquad (15.18)$$

D'après le théorème qui établit la dépendance continue entre les solutions des équations différentielles et les conditions initiales, il existe un voisinage $W^{0\prime} \subset W^0$ du point x^0 tel que toute solution du système d'équations analogue au système (1) (de même commande $u(t)$ $(t_0 \leqslant t < t_1)$)

$$\frac{dy_j}{dt} = f_j(y_1, \ldots, y_n, u_1, \ldots, u_r) \quad (j = 1, \ldots, n), \qquad (15.19)$$

telle que $y(t_0) \in W^{0\prime}$, est définie dans l'intervalle $t_0 \leqslant t \leqslant t_1$ tout entier et vérifie la relation $y(t_1) \in W^*$.

Conformément à la condition du lemme démontré on peut admettre que la solution $y(t)$ $(t_0 \leqslant t \leqslant t_1)$ du système (19) aux données initiales $y(t_0) \in W^{0\prime}$ ne coupe M qu'en un nombre fini de points. En vertu du lemme 2 on en déduit l'inégalité

$$\omega(y(t_1)) - \omega(y(t_0)) \leqslant t_1 - t_0. \qquad (15.20)$$

Ensuite, puisque $y(t_0) \in W^{0\prime} \subset W^0$, il vient $y(t_1) \in W^*$. Par conséquent, d'après (18), on a

$$\left.\begin{array}{l} |\,\omega(y(t_0)) - \omega(x^0)\,| < \varepsilon, \\ |\,\omega(y(t_1)) - \omega(x^*)\,| < \varepsilon. \end{array}\right\} \qquad (15.21)$$

Les relations (21) entraînent

$$\left.\begin{array}{r} \omega(y(t_0)) - \omega(x^0) < \varepsilon, \\ -\omega(y(t_1)) + \omega(x^*) < \varepsilon. \end{array}\right\} \qquad (15.22)$$

En additionnant les premiers et les seconds membres des inégalités (20) et (22), on obtient

$$\omega(x^*) - \omega(x^0) < t_1 - t_0 + 2\varepsilon. \qquad (15.23)$$

Comme ε est arbitraire, on en tire l'inégalité (14).

Le lemme 3 donne les conditions les plus générales pour que le critère (14) soit vrai. Toutefois, ces conditions sont formulées séparément pour chaque commande $u(t)$. Il reste encore à montrer que ces conditions sont vérifiées pour toute commande $u(t)$.

Dans les ouvrages déjà cités, V. Boltianski montre que sous certaines contraintes imposées à l'ensemble M les conditions du lemme 3 sont respectées quelles que soient les commandes admissibles $u(t)$. Il en est ainsi, en particulier, si M est un ensemble lisse par morceaux de dimension $<n$.

Donnons la définition de l'ensemble lisse par morceaux [15]. Soit K un polyèdre convexe borné de dimension s ($s \leqslant n$) qui repose dans l'espace vectoriel Ξ des variables $\xi_1, \xi_2, \ldots, \xi_s$ et qui est considéré avec sa frontière (c'est-à-dire qu'il est fermé). Supposons que dans un ensemble ouvert de l'espace Ξ qui contient le polyèdre K sont données n fonctions continûment dérivables

$$\varphi_i (\xi_1, \xi_2, \ldots, \xi_s) \qquad (i = 1, 2, \ldots, n), \tag{15.24}$$

qui jouissent de la propriété suivant laquelle le rang de la matrice fonctionnelle

$$\begin{bmatrix} \frac{\partial\varphi_1}{\partial\xi_1} & \frac{\partial\varphi_2}{\partial\xi_1} & \cdots & \frac{\partial\varphi_n}{\partial\xi_1} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial\varphi_1}{\partial\xi_s} & \frac{\partial\varphi_2}{\partial\xi_s} & \cdots & \frac{\partial\varphi_n}{\partial\xi_s} \end{bmatrix}$$

en tout point $\xi \in K$ est s. Les fonctions (24) réalisent une application lisse φ du polyèdre K dans l'espace X d'après les formules

$$x_i = \varphi_i (\xi_1, \xi_2, \ldots, \xi_s) \qquad (i = 1, \ldots, n). \tag{15.25}$$

Si cette application est biunivoque (c'est-à-dire si elle associe aux points distincts du polyèdre K des points distincts de l'espace X), l'image $L = \varphi (K)$ du polyèdre K est dite *polyèdre curviligne de dimension s dans l'espace X*. Il est clair que le polyèdre curviligne est un ensemble borné et fermé de l'espace X.

Soit maintenant D un ensemble ouvert de l'espace de phase X. Tout ensemble $M \subset D$ qui est une réunion d'un nombre fini ou infini de polyèdres curvilignes tels que tout ensemble borné fermé reposant dans D ne coupe qu'un nombre fini de ces polyèdres, s'appelle *ensemble lisse par morceaux* dans D. (Les polyèdres peuvent s'accumuler vers la frontière de D.)

Si parmi les polyèdres curvilignes dont la réunion est un ensemble lisse par morceaux M il y a au moins un polyèdre de dimension k, alors que tout autre polyèdre est de dimension inférieure ou égale à k, l'ensemble lisse par morceaux est dit de dimension k. En particulier, toute surface lisse fermée dans D de dimension inférieure à n est un ensemble lisse par morceaux dans D (du fait qu'elle peut être partitionnée en polyèdres curvilignes). Ce qui vient d'être dit entraîne qu'un ensemble lisse par morceaux dans D de dimension inférieure à n ne possède pas de points intérieurs.

Indiquons à titre d'exemple que pour $n = 2$, les ensembles lisses par morceaux de dimension <2 sont composés de lignes formées de morceaux dérivables particuliers (ligne AOB de la fig. 15.1, par exemple).

Le lemme suivant est donné sans démonstration du fait que celle-ci impose l'utilisation des notions relatives à la théorie des variétés lisses; cette démonstration est exposée dans les ouvrages cités de V. Boltianski [15].

L e m m e 4. *Soit M un ensemble lisse par morceaux dans D de dimension $\leqslant n - 1$. Soit, ensuite, $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ une commande admissible qui ramène le point représentatif de la position x^0 en position x^*; de plus, la trajectoire correspondante $x(t)$ $(t_0 \leqslant t \leqslant t_1)$ repose tout entière dans l'ensemble D. Alors, dans tout voisinage W^0 du point x^0 il existe un point y^0 tel que la trajectoire $y(t)$ $(t_0 \leqslant t \leqslant t_1)$ issue du point y^0 et associée à la commande $u(t)$, repose toute entière dans D*

et ne coupe M qu'en un nombre fini de points, c'est-à-dire il n'existe qu'un nombre fini d'instants t ($t_0 \leqslant t \leqslant t_1$) *tels que* $y(t) \in M$.

Les lemmes 3 et 4 entraînent immédiatement le lemme principal suivant.

Lemme principal. *Soit M un ensemble lisse par morceaux de dimension* $<n$ *contenu dans l'ensemble ouvert D de l'espace de phase X. Supposons qu'on donne sur l'ensemble D une fonction continue* $\omega(x) = \omega(x_1, \ldots, x_n)$ *telle qu'aux points n'appartenant pas à l'ensemble M, elle possède des dérivées continues et vérifie l'inégalité* (8)

$$\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \leqslant 1.$$

Alors, si $u(t)$ ($t_0 \leqslant t \leqslant t_1$) *est une commande admissible qui ramène le point représentatif de la position* x^0 *en position* x^*, *la trajectoire correspondante* $x(t)$ *reposant tout entière dans l'ensemble D, le critère* (14)

$$t_1 - t_0 \geqslant \omega(x^*) - \omega(x^0)$$

est vrai pour la durée de la transition.

Les résultats exposés ci-dessus conduisent au théorème suivant, obtenu dans l'ouvrage [14] de V. Boltianski, qui exprime la condition nécessaire et suffisante d'optimalité sous la forme de la méthode de programmation dynamique pour le problème de la réponse en temps minimal.

Théorème [14]. *Soit M un ensemble lisse par morceaux de dimension* $<n$ *inclus dans l'espace de phase X, et* $\omega(x) = \omega(x_1, \ldots, x_n)$, *une fonction continue donnée sur X et possédant des dérivées continues aux points qui n'appartiennent pas à l'ensemble M. Soit, ensuite* $\omega(x^*) = 0$ *pour un point* $x^* \in X$. *Supposons que pour tout point* $x^0 \in X$ *différent de* x^*, *il existe une commande admissible* $u(t) = u_{x^0}(t)$ *qui ramène le point représentatif de la position* x^0 *en position* x^* *en un temps* $-\omega(x^0)$. *Pour que toute commande* $u_{x^0}(t)$ *soit optimale, il faut et il suffit qu'en tout point x n'appartenant pas à l'ensemble M, la fonction* $\omega(x)$ *vérifie l'équation de Bellman* (7)

$$\max_{u \in \Omega} \left[\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \right] = 1$$

ou, ce qui revient au même, l'inégalité (8)

$$\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \leqslant 1 \qquad (u \in \Omega).$$

Démonstration. La nécessité de la condition du théorème est démontrée de même qu'à la page 226. La démonstration que l'on

y donne est applicable à tout point $x = x^0$ où les dérivées $\frac{\partial \omega}{\partial x_i}$ $(i = 1, \ldots, n)$ existent et sont continues.

La suffisance résulte du lemme principal. En effet, il résulte de ce dernier que si la relation (8) est vérifiée en tout point x n'appartenant pas à l'ensemble M et, de plus, si $\omega(x^*) = 0$, la commande qui ramène le système de x^0 à x^* en un temps $-\omega(x^0)$ est optimale. Pour $\omega(x^*) = 0$, le critère de la durée de transition du système de x^0 en x^*, quelle que soit la commande admissible, est

$$t_1 - t_0 \geqslant -\omega(x^0),$$

c'est-à-dire cette transition ne peut être réalisée plus vite qu'en un temps $-\omega(x^0)$.

Dans le théorème qui vient d'être démontré les contraintes imposées à la fonction $\omega(x)$ d'être continûment dérivable sont si faibles que ce théorème peut être appliqué à un très grand nombre de problèmes variés. C'est ce qui justifie la méthode de programmation dynamique.

3. Conditions suffisantes d'optimalité du problème général de programmation dynamique. Pour ce problème, l'équation de Bellman (14.36) a été obtenue au § 14, point 2. L'expression (14.36) entraîne que pour toute commande $u(t) \in \Omega$ qui fait passer le système de la position initiale $x(t)$ en position x^*, on a l'inégalité

$$G(x, u) + \sum_{i=1}^{n} \frac{\partial \Phi(x)}{\partial x_i} f_i(x, u) \geqslant 0 \tag{15.26}$$

ou

$$\sum_{i=1}^{n} \frac{\partial \Phi(x)}{\partial x_i} f_i(x, u) \geqslant -G(x, u) \quad (u \in \Omega), \tag{15.27}$$

où l'égalité a lieu pour la commande optimale $u^*(t)$.

Maintenant, si à l'aide de la relation

$$\omega(x) = -\Phi(x), \tag{15.28}$$

on introduit la fonction $\omega(x)$, l'inégalité (27) devient

$$\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \leqslant G(x, u) \qquad (u \in \Omega). \tag{15.29}$$

D'une façon analogue au point 2 on démontre pour le problème général traité ici le théorème qui donne les conditions nécessaires et suffisantes d'optimalité [14].

T h é o r è m e. *Soit M un ensemble lisse par morceaux de dimension $< n$ contenu dans l'espace de phase X et $\omega(x) = \omega(x_1, \ldots, x_n)$*

une fonction continue donnée sur X et possédant des dérivées continues aux points qui n'appartiennent pas à M. Soit, ensuite, $\omega(x^*) = 0$ *pour un point* $x^* \in X$. *Supposons que pour tout point* $x^0 \in X$ *différent de* x^*, *il existe une commande admissible* $u(t) = u_{x^0}(t)$ *qui ramène le point représentatif de la position* x^0 *en position* x^* *et qui vérifie la relation*

$$\int_{t_0}^{t_1} G(x(t), u(t))\, dt = -\omega(x^0). \tag{15.30}$$

Pour que toute commande $u_{x^0}(t)$ *soit optimale, il faut et il suffit qu'en tout point* x *n'appartenant pas à l'ensemble* M, *la fonction* $\omega(x)$ *vérifie l'équation de Bellman* (14.36)

$$\min_{u \in \Omega} \left[G(x, u) - \sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \right] = 0$$

ou, ce qui revient au même, l'inégalité (29)

$$\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \leqslant G(x, u) \qquad (u \in \Omega).$$

§ 16. Relation entre l'équation de Bellman et l'équation de Jacobi-Hamilton dans les problèmes de mécanique analytique

1. Problème de minimisation de l'intégrale $Q = \int_{t_0}^{t} G(x, u, v)\, dv$.

Considérons le système décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = f(x, u, t), \tag{16.1}$$

où x, f, ainsi que u sont des vecteurs de dimension n

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_n \end{bmatrix}, \quad f(x, u, t) = \begin{bmatrix} f_1(x, u, t) \\ \cdots\cdots \\ f_n(x, u, t) \end{bmatrix}.$$

L'état initial du système

$$[x(t)]_{t=t_0} = x(t_0) \tag{16.2}$$

et l'état du système à l'instant t sont donnés. Trouver la commande

$$u = u^*(t), \tag{16.3}$$

qui ramène le système de l'état $x(t_0)$ à l'état $x(t)$ en minimisant l'intégrale

$$Q=\int_{t_0}^{t} G(x(\nu),\ u(\nu),\ \nu)\,d\nu. \qquad (16.4)$$

Introduisons les notations

$$W(x(t),\ t)=\min_{u(\nu)}\int_{t_0}^{t} G(x(\nu),\ u(\nu),\ \nu)\,d\nu. \qquad (16.5)$$

Soit $u^*(\nu)$ $(0\leqslant\nu\leqslant t)$ la commande minimisant l'intégrale (4) et $x(\nu)$ la trajectoire de phase (fig. 16.1) associée à cette commande. Il vient

$$W(x(t),\ t)=\int_{t_0}^{t} G(x(\nu),\ u^*(\nu),\ \nu)\,d\nu.$$

Prenons sur la trajectoire optimale (fig. 16.1) le point $x(t-\Delta t)$. Puisque la position du système au point final $t-\Delta t$ de l'intervalle

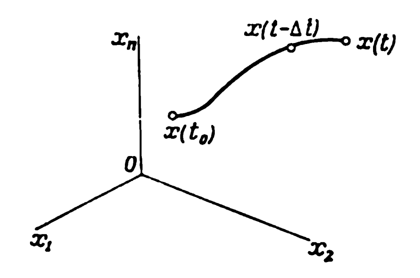


Fig. 16.1

$(t_0,\ t-\Delta t)$ est fixée, l'arc de la trajectoire optimale entre les points $x(t_0)$ et $x(t-\Delta t)$ est lui-même une trajectoire optimale. C'est pourquoi pour $W(x(t-\Delta t),\ t-\Delta t))$ on a l'expression suivante

$$W(x(t-\Delta t),\ t-\Delta t)=W(x(t),\ t)-\int_{t-\Delta t}^{t} G(x(\nu),\ u^*(\nu),\ \nu)\,d\nu. \qquad (16.6)$$

Comme

$$\int_{t-\Delta t}^{t} G(x(\nu),\ u^*(\nu),\ \nu)\,d\nu=G(x(t),\ u^*(t),\ t)\,\Delta t+o_1(\Delta t), \qquad (16.7)$$

on a

$$W(x(t-\Delta t),\ t-\Delta t)=W(x(t),\ t)-G(x(t),\ u^*(t),\ t)\,\Delta t-o_1(\Delta t). \qquad (16.8)$$

Supposons maintenant que la fonction $W(x, t)$ est continue et possède partout des dérivées partielles continues par rapport à tous ses variables. Puisque d'après (1) pour $u = u^*(t)$

$$x(t-\Delta t) = x(t) - \frac{dx}{dt}\Delta t + o_2(\Delta t) =$$
$$= x(t) - f(x(t), u^*(t), t)\Delta t + o_2(t), \qquad (16.9)$$

on a

$$W(x(t-\Delta t), t-\Delta t) = W(x(t) - f(x(t), u^*(t), t)\Delta t +$$
$$+ o_2(\Delta t), t-\Delta t) = W(x(t), t) - \sum_{i=1}^{n} \frac{\partial W(x, t)}{\partial x_i} f_i(x, u^*, t)\Delta t -$$
$$-\frac{\partial W(x, t)}{\partial t}\Delta t + o_3(\Delta t). \qquad (16.10)$$

En portant l'expression (10) dans le premier membre de l'équation (8), on obtient

$$-\sum_{i=1}^{n} \frac{\partial W(x, t)}{\partial x_i} f_i(x, u^*, t)\Delta t - \frac{\partial W(x, t)}{\partial t}\Delta t =$$
$$= -G(x, u^*, t)\Delta t + o_4(\Delta t). \qquad (16.11)$$

En divisant tous les termes de (11) par Δt et en tenant compte que $\lim\limits_{\Delta t\to 0} \frac{o(\Delta t)}{\Delta t} = 0$, on obtient à la limite pour $\Delta t \to 0$ la relation

$$G(x, u^*, t) - \sum_{i=1}^{n} \frac{\partial W(x, t)}{\partial x_i} f_i(x, u^*, t) - \frac{\partial W(x, t)}{\partial t} = 0 \qquad (16.12)$$

ou la relation équivalente

$$G(x, u^*, t) - \sum_{i=1}^{n} \frac{\partial W(x, t)}{\partial x_i} \frac{dx_i}{dt} - \frac{\partial W(x, t)}{\partial t} = 0. \qquad (16.13)$$

Les raisonnements qui nous ont conduits à la relation (6) entraînent que la relation (13) conserve encore sa forme dans le cas du problème de *maximisation* de l'intégrale (4).

2. Obtention de l'équation de Jacobi-Hamilton à partir du principe de Hamilton. Le principe de Hamilton consiste dans le fait que pour un mouvement réel entre deux configurations du système données aux instants t_0 et t (il s'agit de la position du système dans un espace $q_1, q_2, \ldots, q_n$ et non pas dans un espace de phase), la valeur de l'intégrale $\int_{t_0}^{t} L\,dt$ (où L est la fonction de Lagrange) est stationnaire. Désignons cette valeur stationnaire par W.

La valeur stationnaire de l'intégrale $\int_{t_0}^{t} L\,dt$ est assurée en vertu des équations du mouvement du système

$$\left.\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i}, \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i} \qquad (i=1,\ \ldots,\ n), \end{aligned}\right\} \tag{16.14}$$

On peut pourtant adopter un *autre point de vue*.

Supposons que l'espace des configurations $q_1, \ldots, q_n$ du système (14) est l'espace de phase d'un autre système régi par les équations

$$\frac{dq_i}{dt} = f_i(q_1, \ldots, q_n, p_1, \ldots, p_n, t) \qquad (i=1, \ldots, n), \tag{16.15}$$

où

$$f_i(q_1, \ldots, q_n, p_1, \ldots, p_n, t) = \frac{\partial H}{\partial p_i} \tag{16.16}$$

est une fonction connue de ses variables. Ainsi, on connaît la forme des fonctions f_i des seconds membres des équations (15).

Le mouvement réel assure à l'intégrale $\int_{t_0}^{t} L\,dt$ une valeur stationnaire. On peut donc identifier en pensée $p_1, p_2, \ldots, p_n$ à des commandes $u_1^*, u_2^*, \ldots, u_n^*$ du système

$$\frac{dq_i}{dt} = f_i(q_1, \ldots, q_n, u_1, \ldots, u_n, t) \qquad (i=1, \ldots, n), \tag{16.17}$$

telles qu'elles extrémisent l'intégrale $\int_{t_0}^{t} L\,dt$, c'est-à-dire à des commandes optimales du système (17).

D'après ce qui a été démontré, les fonctions $u^*(t)$ intervenant dans la relation (13) sont elles-mêmes des commandes optimales. C'est pourquoi la relation (13) du système décrit par les équations (17) peut être mise sous la forme

$$L(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n, t) = \\ = \sum_{i=1}^{n} \frac{\partial W(q, t)}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial W(q, t)}{\partial t}. \tag{16.18}$$

On sait que pour le système (14) la fonction de Hamilton est

$$\hat{H}(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n, t) = \\ = -L(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n, t) + \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i. \tag{16.19}$$

En tenant compte du fait que

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \qquad (i = 1, \ldots, n), \tag{16.20}$$

et en exprimant $\dot{q}_i$ par les variables canoniques

$$\dot{q}_i = \dot{q}_i (q_1, \ldots, q_n, p_1, \ldots, p_n, t) \qquad (i = 1, \ldots, n), \tag{16.21}$$

on ramène la fonction $\hat{H}$ à des variables canoniques

$$H = H (q_1, \ldots, q_n, p_1, \ldots, p_n, t). \tag{16.22}$$

L'équation (18) peut s'écrire

$$\frac{\partial W}{\partial t} - L (q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n, t) + \sum_{i=1}^{n} \frac{\partial W}{\partial q_i} \dot{q}_i = 0. \tag{16.23}$$

En dérivant par rapport à $\dot{q}_i$ le premier et le second membre de la relation (18), on trouve

$$\frac{\partial L}{\partial \dot{q}_i} = \frac{\partial W (q, t)}{\partial q_i} \qquad (i = 1, \ldots, n).$$

Il s'ensuit qu'en remplaçant dans les expressions (21) les variables p_i par

$$p_i = \frac{\partial W}{\partial q_i} \qquad (i = 1, \ldots, n), \tag{16.24}$$

on obtient

$$\dot{q}_i = \dot{q}_i \left(q_1, \ldots, q_n, \frac{\partial W}{\partial q_1}, \ldots, \frac{\partial W}{\partial q_n}, t \right) \qquad (i = 1, \ldots, n). \tag{16.25}$$

En substituant maintenant dans l'équation (23) à $\dot{q}_i$ les expressions (25) on ramène (23) à la forme

$$\frac{\partial W}{\partial t} + H \left(q_1, \ldots, q_n, \frac{\partial W}{\partial q_1}, \ldots, \frac{\partial W}{\partial q_n}, t \right) = 0. \tag{16.26}$$

L'équation aux dérivées partielles (26) est précisément l'équation de Hamilton-Jacobi pour un système mécanique dont le mouvement est décrit par les équations différentielles (14).

Exemple. Considérons à titre d'exemple le système à un degré de liberté où l'énergie cinétique T et l'énergie potentielle V s'écrivent

$$T = \frac{1}{2} m\dot{q}^2, \qquad V = \frac{1}{2} cq^2. \tag{16.27}$$

La fonction de Lagrange est

$$L = \frac{1}{2} m\dot{q}^2 - \frac{1}{2} cq^2. \tag{16.28}$$

Comme

$$p = \frac{\partial L}{\partial \dot{q}} = m\dot{q}, \tag{16.29}$$

on a

$$\dot{q}=\frac{p}{m}. \tag{16.30}$$

D'après (19), dans le problème traité la fonction de Hamilton s'écrit

$$\hat{H}=-L+\frac{\partial L}{\partial \dot{q}}\dot{q}=\frac{m\dot{q}^2}{2}+\frac{cq^2}{2}. \tag{16.31}$$

En remplaçant dans (31) $\dot{q}$ par l'expression (30), on ramène la fonction de Hamilton aux variables canoniques

$$H=\frac{p^2}{2m}+\frac{cq^2}{2}. \tag{16.32}$$

Les équations canoniques du mouvement du système

$$\left.\begin{aligned}\frac{dq}{dt}&=\frac{\partial H}{\partial p},\\ \frac{dp}{dt}&=-\frac{\partial H}{\partial q}\end{aligned}\right\} \tag{16.33}$$

deviennent en vertu de (32)

$$\left.\begin{aligned}\frac{dq}{dt}&=\frac{1}{m}p,\\ \frac{dp}{dt}&=-cq.\end{aligned}\right\} \tag{16.34}$$

Dans l'exemple considéré, *l'action hamiltonienne* est la fonction

$$W=\int_{t_0}^{t} L\,dt=\int_{t_0}^{t}\left[\frac{1}{2}m\dot{q}^2(t)-\frac{1}{2}cq^2(t)\right]dt. \tag{16.35}$$

D'après (34) et (30) l'équation (17) de notre exemple se met sous la forme

$$\frac{dq}{dt}=\frac{1}{m}u. \tag{16.36}$$

Conformément à (28) la fonctionnelle (4) à minimiser est ici

$$Q=\int_{t_0}^{t}\left[\frac{u^2(t)}{2m}-\frac{cq^2(t)}{2}\right]dt. \tag{16.37}$$

La valeur extrémale W de la fonctionnelle Q est atteinte lorsque la commande optimale $u=u^*(t)$ qui en vertu du principe de Hamilton doit être

$$u^*(t)=p, \tag{16.38}$$

car ce n'est que pour une telle commande que l'action W satisfait à la condition $\delta W=0$.

L'équation (13) devient alors

$$\frac{p^2}{2m}-\frac{1}{2}cq^2-\frac{\partial W}{\partial q}\dot{q}-\frac{\partial W}{\partial t}=0. \tag{16.39}$$

D'après (24), remplaçons dans l'expression (30) la variable p par

$$p=\frac{\partial W}{\partial q}. \tag{16.40}$$

Après cette substitution l'expression (30) devient

$$\dot{q}=\frac{1}{m}\frac{\partial W}{\partial q}. \tag{16.41}$$

En portant dans (39) au lieu de $\dot{q}$ l'expression (41), ramenons l'équation (39) à la forme

$$\frac{\partial W}{\partial t}+\frac{1}{2m}\left(\frac{\partial W}{\partial q}\right)^2+\frac{1}{2}cq^2=0. \tag{16.42}$$

L'équation (42) est l'équation de Jacobi-Hamilton relative à l'exemple considéré.

Comme d'après (32) la fonction de Hamilton de l'exemple traité est de la forme

$$H(q,\ p,\ t)=\frac{p^2}{2m}+\frac{cq^2}{2},$$

l'équation (42) peut s'écrire

$$\frac{\partial W}{\partial t}+H\left(q,\ \frac{\partial W}{\partial q},\ t\right)=0. \tag{16.43}$$

CHAPITRE 5

PRINCIPE DU MAXIMUM DE L. PONTRIAGUINE DANS LA THÉORIE DE LA COMMANDE OPTIMALE

§ 17. Théorème de la condition nécessaire d'optimalité

1. Principe du maximum de Pontriaguine. Le système commandé est décrit par les équations différentielles

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r) \qquad (j = 1, \ldots, n). \tag{17.1}$$

On demande de ramener le système du point $x(t_0)$ de l'espace de phase X de dimension n au point donné x^*. L'instant t_1 où le point représentatif tombe en x^* n'est pas fixé à l'avance.

La commande $u(t)$ est un vecteur fonction continu par morceaux dont les valeurs appartiennent à un certain ensemble borné fermé Ω de l'espace U de dimension r

$$u \in \Omega. \tag{17.2}$$

Les fonctions $f_j(x_1, \ldots, x_n, u_1, \ldots, u_r)$ $(j = 0, 1, \ldots, n)$ sont définies pour des valeurs quelconques de $x \in X$ et $u \in \Omega$. On les suppose continues par rapport à l'ensemble des variables $x_1, \ldots, x_n, u_1, \ldots, u_r$ et continûment dérivables par rapport à $x_1, \ldots, x_n$.

La commande u doit être choisie de façon à minimiser la fonctionnelle

$$Q = \int_{t_0}^{t_1} f_0(x(\xi), u(\xi))\, d\xi. \tag{17.3}$$

Désignons par $x_0(t)$ la fonction déterminée par l'équation différentielle

$$\frac{dx_0}{dt} = f_0(x_1, \ldots, x_n, u_1, \ldots, u_r) \tag{17.4}$$

et par la condition initiale

$$x_0(t_0) = 0. \tag{17.5}$$

La fonctionnelle Q à minimiser peut être mise sous la forme

$$Q = x_0(t_1). \tag{17.6}$$

16—0393

Supposons qu'en plus du système d'équations principal (1) et de l'équation (4) qu'on peut écrire ensemble

$$\frac{dx_j}{dt} = f_j(x, u) \qquad (j = 0, 1, \ldots, n), \tag{17.7}$$

on ait encore un système d'équations par rapport aux variables auxiliaires $\psi_0, \psi_1, \ldots, \psi_n$:

$$\frac{d\psi_k}{dt} = -\sum_{i=0}^{n} \frac{\partial f_i(x, u)}{\partial x_k} \psi_i \qquad (k = 0, 1, \ldots, n). \tag{17.8}$$

Si nous choisissons une commande admissible $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ qui fait passer le système (1) du point $x(t_0)$ en x^*, et si nous avons une trajectoire de phase correspondante $x(t)$ du système (7) à condition initiale $x(t_0)$, $x_0(t_0) = 0$, le système d'équations (8) devient

$$\frac{d\psi_k}{dt} = -\sum_{i=0}^{n} \frac{\partial f_i(x(t), u(t))}{\partial x_k} \psi_i \qquad (k = 0, 1, \ldots, n). \tag{17.9}$$

C'est un système des équations différentielles linéaires homogènes à coefficients variables. Quelles que soient les valeurs initiales de ψ_i il admet une seule solution

$$\psi = (\psi_0, \psi_1, \ldots, \psi_n).$$

Toute solution du système d'équations (9) (quelles que soient les conditions initiales) est dite solution du système (8) qui correspond à la commande retenue $u(t)$ et à la trajectoire de phase $x(t)$.

En introduisant les notations

$$\widetilde{H}(\psi, x, u) = \sum_{i=0}^{n} \psi_i f_i(x, u), \tag{17.10}$$

on aura

$$\frac{\partial \widetilde{H}}{\partial \psi_j} = f_j(x, u), \tag{17.11}$$

$$\frac{\partial \widetilde{H}}{\partial x_k} = \sum_{i=0}^{n} \psi_i \frac{\partial f_i(x, u)}{\partial x_k}, \tag{17.12}$$

et les systèmes d'équations (7) et (8) peuvent s'écrire

$$\frac{dx_j}{dt} = \frac{\partial \widetilde{H}}{\partial \psi_j} \qquad (j = 0, 1, \ldots, n), \tag{17.13}$$

$$\frac{d\psi_j}{dt} = -\frac{\partial \widetilde{H}}{\partial x_j} \qquad (j = 0, 1, \ldots, n). \tag{17.14}$$

Le problème posé ci-dessus se résout à l'aide du principe du maximum de Pontriaguine énoncé et démontré [72] sous la forme du théorème qui suit.

Faisons au préalable les remarques suivantes. En prenant une commande admissible arbitraire $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ (c'est-à-dire une commande qui ramène le système du point $x(t_0)$ au point x^* et qui satisfait aux contraintes $u \in \Omega$) et la condition initiale $x(t_0)$, $x_0(t_0) = 0$, on peut trouver la trajectoire $x(t) = (x_0(t), x_1(t), \ldots$ $\ldots, x_n(t))$ correspondante (c'est-à-dire vérifiant le système (13)). Ensuite on peut obtenir les solutions du système (14) associées aux fonctions $u(t)$ et $x(t)$

$$\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t)).$$

Lorsque les valeurs de ψ et de x sont fixées (constantes), la fonction $\widetilde{H}$ des variables $x_1, \ldots, x_n, \psi_0, \psi_1, \ldots, \psi_n, u_1, \ldots, u_r$

$$\widetilde{H}(\psi, x, u) = \sum_{i=0}^{n} \psi_i f_i(x, u)$$

devient fonction du paramètre $u \in \Omega$; désignons la borne supérieure des valeurs de cette fonction par $\widetilde{M}(\psi, x)$

$$\widetilde{M}(\psi, x) = \sup_{u \in \Omega} \widetilde{H}(\psi, x, u). \tag{17.15}$$

Si la borne supérieure exacte des valeurs de la fonction continue $\widetilde{H}$ est atteinte en un certain point du domaine de commande Ω, $\widetilde{M}(\psi, x)$ est alors le *maximum* des valeurs de la fonction $\widetilde{H}$ pour ψ et x fixées. Par suite le théorème énoncé ci-dessous (condition nécessaire d'optimalité), dont le contenu principal est traduit par l'égalité (16), est nommé par son auteur *principe du maximum*.

Théorème 1 (principe du maximum de L. Pontriaguine [72]). *Soit $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ une commande admissible telle que la trajectoire $x(t)$ qui lui correspond, issue à l'instant t_0 du point $x(t_0)$, passe à l'instant t_1 par le point x^*. Pour que la commande $u(t)$ et la trajectoire $x(t)$ soient optimales, il faut qu'il existe un vecteur fonction non nul et continu $\psi(t) = (\psi_0(t), \psi_1(t), \ldots$ $\ldots, \psi_n(t))$ associé aux fonctions $u(t)$ et $x(t)$ tel que*

1°. *Pour tout instant t $(t_0 \leqslant t \leqslant t_1)$ qui est un point de continuité de la commande $u(t)$, la fonction $\widetilde{H}(\psi(t), x(t), u)$ de la variable $u \in \Omega$ atteint au point $u = u(t)$ le maximum*

$$\widetilde{H}(\psi(t), x(t), u(t)) = \widetilde{M}(\psi(t), x(t)). \tag{17.16}$$

2°. *A l'instant final t_1 on a les relations*

$$\psi_0(t_1) \leqslant 0, \qquad \widetilde{M}(\psi(t_1), x(t_1)) = 0. \tag{17.17}$$

Il s'avère ensuite que si les grandeurs $\psi(t)$, $x(t)$ *et* $u(t)$ *vérifient* *système* (13), (14) *et la condition* 1°, *les fonctions* $\psi_0(t)$ *et* $\tilde{M}(\psi(t), x(t))$ *de la variable* t *sont constantes de façon que la vérification des relations* (17) *peut s'effectuer non pas à l'instant* t_1 *mais à un instant quelconque* t $(t_0 \leqslant t \leqslant t_1)$.

2. Principe du maximum de Pontriaguine dans le problème de la réponse en temps minimal. On peut tirer du théorème 1 la condition nécessaire analogue d'optimalité du problème de la réponse en temps minimal. A cet effet, il faut poser en vertu de (3)

$$f_0(x, u) \equiv 1. \tag{17.18}$$

D'après (10) et (18) la fonction $\tilde{H}$ s'écrit

$$\tilde{H}(\psi, x, u) = \psi_0 + \sum_{i=1}^{n} \psi_i f_i(x, u). \tag{17.19}$$

Désignons par ψ le vecteur de dimension n

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \cdots \\ \psi_n \end{bmatrix}, \tag{17.20}$$

et par $H(\psi, x, u)$ la fonction

$$H(\psi, x, u) = \sum_{i=1}^{n} \psi_i f_i(x, u). \tag{17.21}$$

Les équations (1) et les équations (8) peuvent être mises sous la forme

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial \psi_j} \quad (j = 1, \ldots, n), \tag{17.22}$$

$$\frac{d\psi_j}{dt} = -\frac{\partial H}{\partial x_j} \quad (j = 1, \ldots, n). \tag{17.23}$$

Pour des valeurs fixées de ψ et de x, la fonction H devient fonction du paramètre u; désignons par $M(\psi, x)$ la borne supérieure des valeurs de cette fonction

$$M(\psi, x) = \sup_{u \in \Omega} H(\psi, x, u). \tag{17.24}$$

Puisque d'après (21) et (19)

$$H(\psi, x, u) = \tilde{H}(\psi, x, u) - \psi_0 \tag{17.25}$$

on obtient que

$$M(\psi, x) = \tilde{M}(\psi, x) - \psi_0. \tag{17.26}$$

C'est pourquoi les conditions (16) et (17) s'écrivent

$$H(\psi(t), x(t), u(t)) = M(\psi(t), x(t)) = -\psi_0 \geqslant 0. \qquad (17.27)$$

Ainsi, on obtient le théorème suivant [72].

Théorème 2 (principe du maximum dans le problème de la réponse en temps minimal). *Soit $u(t)$ ($t_0 \leqslant t \leqslant t_1$) une commande admissible qui ramène le point représentatif de $x(t_0)$ en x^*, et $x(t)$, la trajectoire correspondante telle que $x(t_1) = x^*$. Pour que la commande $u(t)$ et la trajectoire $x(t)$ soient optimales (quant à la réponse), il faut qu'il existe un vecteur fonction continu non nul $\psi(t)=(\psi_1(t), \psi_2(t), \ldots, \psi_n(t))$, associé aux fonctions $u(t)$ et $x(t)$, tel que*

1°. *Pour tout instant t ($t_0 \leqslant t \leqslant t_1$), qui est un point de continuité de la commande $u(t)$, la fonction $H(\psi(t), x(t), u)$ de la variable $u \in \Omega$ atteint en $u = u(t)$ le maximum*

$$H(\psi(t), x(t), u(t)) = M(\psi(t), x(t)). \qquad (17.28)$$

2°. *A l'instant final t_1 la relation*

$$M(\psi(t_1), x(t_1)) \geqslant 0 \qquad (17.29)$$

est observée.

Il se trouve ensuite que si les grandeurs $\psi(t)$, $x(t)$, $u(t)$ satisfont au système (22), (23) *et à la condition* 1°, *la fonction $M(\psi(t), x(t))$ de la variable t est constante et il n'est pas de rigueur de vérifier la relation* (29) *à l'instant t_1, on peut le faire à un instant quelconque t ($t_0 \leqslant t \leqslant t_1$).*

3. Démonstration du théorème sur la condition nécessaire d'optimalité (principe du maximum) du problème à temps fixé T et à extrémité libre de la trajectoire. Dans le cas général, la démonstration du principe du maximum est assez compliquée. C'est pourquoi nous allons nous borner à l'exposé d'un seul cas particulier relativement simple.

Considérons le système régi par les équations différentielles

$$\frac{dx_j}{dt} = f_j(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_r) = f_j(x, u) \qquad (17.30)$$
$$(j = 1, \ldots, n),$$

auxquelles est équivalente l'équation différentielle vectorielle

$$\frac{dx}{dt} = f(x, u), \qquad (17.31)$$

où x, u et f sont les vecteurs

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u) = \begin{bmatrix} f_1(x, u) \\ \cdots\cdots \\ f_n(x, u) \end{bmatrix}. \qquad (17.32)$$

La commande $u(t)$ est un vecteur fonction continu par morceaux; elle doit satisfaire aux contraintes

$$u \in \Omega. \tag{17.33}$$

L'état initial du système est donné

$$[x(t)]_{t=t_0} = x(t_0). \tag{17.34}$$

On demande de trouver la commande $u(t)$ ($t_0 \leqslant t \leqslant T$) qui vérifie les contraintes (33) et minimise la fonctionnelle

$$Q = \int_{t_0}^{T} f_0(x(\xi), u(\xi))\, d\xi, \tag{17.35}$$

où T est une grandeur *fixée*.

Désignons par $x_0(t)$, de même qu'au point 1, la fonction scalaire déterminée par l'équation différentielle

$$\frac{dx_0}{dt} = f_0(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_r) = f_0(x, u) \tag{17.36}$$

et par la condition initiale

$$[x_0(t)]_{t=t_0} = 0. \tag{17.37}$$

(36), (37) et (35) entraînent que

$$Q = x_0(T). \tag{17.38}$$

Désignons ici la commande optimale par $\hat{u}(t)$. La trajectoire $\hat{x}(t)$ associée à cette commande est optimale.

La commande optimale $\hat{u}(t)$ est un vecteur dont les coordonnées $\hat{u}_i(t)$ peuvent, comme nous l'avons dit, avoir les discontinuités de première espèce en un nombre fini de points de l'intervalle $t_0 < t < T$.

Examinons un intervalle de temps infiniment petit $\tau - \varepsilon < t < \tau$, où ε est un infiniment petit, et $\tau \in (t_0, T)$.

Attribuons à la commande $\hat{u}(t)$ une variation en remplaçant $\hat{u}(t)$ dans l'intervalle de temps infiniment petit $\tau - \varepsilon < t < \tau$ par une autre commande u (sans changer la commande $\hat{u}(t)$ en dehors de cet intervalle infiniment petit).

Constatons qu'ici on ne demande pas que l'accroissement $u_i(t) - \hat{u}_i(t)$ ($i = 1, \ldots, r$), où $t \in (\tau - \varepsilon, \tau)$, soit infiniment petit. Si la contrainte (33) est, par exemple, de la forme $|u_i| \leqslant m_i$, l'accroissement $u_i(t) - \hat{u}_i(t)$ ($t \in (\tau - \varepsilon, \tau)$) peut prendre n'importe quelle valeur k_i, où $0 \leqslant |k_i| \leqslant 2m_i$.

Toutefois, comme ε est un infiniment petit, l'impulsion $[u_i(t) - \hat{u}_i(t)]\,\varepsilon$ ($i = 1, \ldots, r$) de l'accroissement de la commande est également infiniment petite. Cherchons la variation de la trajectoire du système produite par cette impulsion. En vertu des équations

(30) et (36) on a

$$x_j(\tau)-\hat{x}_j(\tau)=\left[\left(\frac{dx_j}{dt}\right)_{t=\tau}-\left(\frac{d\hat{x}_j}{dt}\right)_{t=\tau}\right]\varepsilon+o(\varepsilon)=$$

$$=[f_j(x(\tau),\ u(\tau))-f_j(\hat{x}(\tau),\ \hat{u}(\tau))]\,\varepsilon+o(\varepsilon) \qquad (17.39)$$

$$(j=0,\ 1,\ \ldots,\ n).$$

D'après (39), la différence $x_j(\tau)-\hat{x}_j(\tau)$ est une grandeur de même ordre de petitesse que ε. Il s'ensuit que $f_j(x(\tau),\ u(\tau))-$ $-f_j(\hat{x}(\tau),\ u(\tau))$ est également une grandeur de même ordre de petitesse que ε. Il en résulte que la relation (39) peut être remplacée par la relation

$$x_j(\tau)-\hat{x}_j(\tau)=[f_j(\hat{x}(\tau),\ u(\tau))-f_j(\hat{x}(\tau),\ \hat{u}(\tau))]\,\varepsilon+o_1(\varepsilon) \qquad (17.40)$$

$$(j=0,\ 1,\ \ldots,\ n).$$

Introduisons maintenant les fonctions $\delta x_j(t)$ $(j=0,\ 1,\ \ldots,\ n)$ à l'aide des relations

$$x_j(t)=\hat{x}_j(t)+\delta x_j(t) \qquad (\tau\leqslant t\leqslant T;\ j=0,\ 1,\ 2,\ \ldots,\ n). \qquad (17.41)$$

Les fonctions $\delta x_j(t)$ s'appellent variations des coordonnées.

Comme on le voit de (40), l'ordre de petitesse des variations δx_j est le même que celui de ε. Par suite lors de la recherche des équations auxquelles satisfont les fonctions δx_j $(j=0,\ 1,\ \ldots,\ n)$ nous négligerons les grandeurs $o(\varepsilon)$.

D'après (40) et (41), à l'instant $t=\tau$ les valeurs des variations δx_j sont

$$\delta x_j(\tau)=[f_j(\hat{x}(\tau),\ u(\tau))-f_j(\hat{x}(\tau),\ \hat{u}(\tau))]\,\varepsilon. \qquad (17.42)$$

Dans l'intervalle de temps $\tau\leqslant t\leqslant T$, la commande u coïncide avec la commande optimale $\hat{u}(t)$. Les équations différentielles vérifiées par les variations $\delta x_j(t)$ $(\tau\leqslant t\leqslant T)$ peuvent s'obtenir de la façon suivante. D'après (30), (36) et (41), on a

$$\frac{dx_j}{dt}+\frac{d(\delta x_j)}{dt}=f_j(\hat{x}_1+\delta x_1,\ \ldots,\ \hat{x}_n+\delta x_n,\ \hat{u}_1,\ \ldots,\ \hat{u}_r) \qquad (17.43)$$

$$(j=0,\ 1,\ \ldots,\ n).$$

En développant les seconds membres de (43) en séries de Taylor dans le voisinage de $\hat{x}_1(t),\ \ldots,\ \hat{x}_n(t)$, on obtient

$$\frac{d\hat{x}_j}{dt}+\frac{d(\delta x_j)}{dt}=f_j(\hat{x}_1,\ \ldots,\ \hat{x}_n,\ \hat{u}_1,\ \ldots,\ \hat{u}_r)+$$

$$+\sum_{k=0}^{n}\frac{\partial f_j(\hat{x}_1,\ \ldots,\ \hat{x}_n,\ \hat{u}_1,\ \ldots,\ \hat{u}_r)}{\partial x_k}\,\delta x_k+o(\delta x_0,\ \delta x_1,\ \ldots,\ \delta x_n)$$

$$(j=0,\ 1,\ \ldots,\ n). \qquad (17.44)$$

D'après (30) et (36), les premiers termes des premiers et des seconds membres des équations (44) s'annulent. C'est pourquoi en rejetant dans (44) les termes qui contiennent δx_i à la puissance deux et aux puissances supérieures et dont l'ensemble se note o $(\delta x_0, \delta x_1, \ldots \ldots, \delta x_n)$, on obtient le système d'équations différentielles linéaires suivant par rapport à δx_j

$$\frac{d(\delta x_j)}{dt} = \sum_{k=0}^{n} \frac{\partial f_j(\hat{x}, \hat{u})}{\partial x_k} \delta x_k \qquad (j = 0, 1, 2, \ldots, n). \qquad (17.45)$$

Les équations différentielles (45) auxquelles satisfont les variations δx_j s'appellent *équations aux variations*.

Notons que si l'on introduit les matrices

$$F(x, u) = \begin{bmatrix} \frac{\partial f_0}{\partial x_0} & \frac{\partial f_0}{\partial x_1} & \frac{\partial f_0}{\partial x_2} & \ldots & \frac{\partial f_0}{\partial x_n} \\ \frac{\partial f_1}{\partial x_0} & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \ldots & \frac{\partial f_1}{\partial x_n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{\partial f_n}{\partial x_0} & \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \ldots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}, \quad \delta x = \begin{bmatrix} \delta x_0 \\ \delta x_1 \\ \ldots \\ \delta x_n \end{bmatrix}, \qquad (17.46)$$

le système d'équations différentielles scalaires (45) peut être mis sous la forme d'une équation différentielle vectorielle équivalente

$$\frac{d(\delta x)}{dt} = F(\hat{x}, \hat{u})\, \delta x. \qquad (17.47)$$

La matrice $F^*(x, u)$ transposée de la matrice $F(x, u)$ a, d'après (46), la forme

$$F^*(x, u) = \begin{bmatrix} \frac{\partial f_0}{\partial x_0} & \frac{\partial f_1}{\partial x_0} & \frac{\partial f_2}{\partial x_0} & \ldots & \frac{\partial f_n}{\partial x_0} \\ \frac{\partial f_0}{\partial x_1} & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \ldots & \frac{\partial f_n}{\partial x_1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{\partial f_0}{\partial x_n} & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \ldots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}. \qquad (17.48)$$

Désignons par ψ le vecteur de dimension $(n + 1)$

$$\psi = \begin{bmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \ldots \\ \psi_n \end{bmatrix}, \qquad (17.49)$$

qui vérifie l'équation différentielle vectorielle

$$\frac{d\psi}{dt} = -F^*(\hat{x}, \hat{u})\,\psi. \tag{17.50}$$

L'équation différentielle vectorielle (50) est équivalente au système d'équations différentielles scalaires

$$\frac{d\psi_k}{dt} = -\sum_{i=0}^{n} \frac{\partial f_i(\hat{x}, \hat{u})}{\partial x_k}\,\psi_i \qquad (k=0,\ 1,\ 2,\ \ldots,\ n). \tag{17.51}$$

Le système d'équations différentielles (51) coïncide avec le système d'équations (9) qui a été introduit au point 1.

Conformément à (47) et à (50), on a la relation

$$\frac{d}{dt}\langle\psi(t),\ \delta x(t)\rangle = \left\langle\frac{d\psi}{dt},\ \delta x\right\rangle + \left\langle\psi,\ \frac{d}{dt}\,\delta x\right\rangle =$$
$$= \langle -F^*\psi,\ \delta x\rangle + \langle\psi,\ F\delta x\rangle = -\psi^*F\delta x + \psi^*F\delta x = 0, \tag{17.52}$$

d'où l'on tire que

$$\sum_{i=0}^{n} \psi_i(t)\,\delta x_i(t) = \text{const} \qquad (\tau \leqslant t \leqslant T). \tag{17.53}$$

L'expression (53) que nous venons d'obtenir est la relation connue de Lagrange pour des systèmes *conjugués* d'équations différentielles linéaires homogènes que sont, par définition, les systèmes représentés par les équations vectorielles (47) et (50).

D'après (38),

$$\delta Q = \delta x_0(T),$$

et comme la commande optimale $\hat{u}(t)$ minimise la fonctionnelle Q, toute autre commande différente de la commande optimale conduit à l'inégalité

$$\delta Q = \delta x_0(T) \geqslant 0. \tag{17.54}$$

Imposons maintenant au système (51) les conditions suivantes qui doivent être vérifiées par la solution de (51) à l'instant $t = T$ (cf. p. 262)

$$\psi_0(T) = -1,\ \psi_1(T) = \psi_2(T) = \ldots = \psi_n(T) = 0. \tag{17.55}$$

On a, en outre, d'après (54) et (55),

$$\sum_{i=0}^{n} \psi_i(T)\,\delta x_i(T) = -\delta x_0(T) = -\delta Q \leqslant 0. \tag{17.56}$$

Grâce à (53), l'inégalité (56) peut s'écrire

$$-\delta Q = \sum_{i=0}^{n} \psi_i(\tau)\,\delta x_i(\tau) \leqslant 0. \tag{17.57}$$

Ici

$$\psi_i(\tau) = [\psi_i(t)]_{t=\tau} \qquad (i = 0, 1, \ldots, n),$$

où $\psi(t)$ est la solution du système (51), qui satisfait aux conditions (55).

Les valeurs de $\delta x_i(\tau)$ $(i = 1, \ldots, n)$ sont connues (elles sont déterminées plus haut par les expressions (42)). En portant ces expressions dans l'inégalité (57), on obtient

$$\sum_{i=0}^{n} \psi_i(\tau)\,[f_i(\hat{x}(\tau), u(\tau)) - f_i(\hat{x}(\tau), \hat{u}(\tau))]\,\varepsilon \leqslant 0. \qquad (17.58)$$

L'intervalle de temps $(\tau - \varepsilon, \tau)$ est celui où la commande optimale $\hat{u}(t)$ a été remplacée par une autre commande. En retenant que $\varepsilon > 0$, l'inégalité (58) peut être remplacée par l'inégalité

$$\sum_{i=0}^{n} \psi_i(\tau)\,[f_i(\hat{x}(\tau), u(\tau)) - f_i(\hat{x}(\tau), \hat{u}(\tau))] \leqslant 0. \qquad (17.59)$$

Puisqu'on peut choisir comme τ un instant courant quelconque t $(t_0 + \varepsilon \leqslant t \leqslant T)$, où ε est une grandeur aussi petite que l'on veut, l'inégalité (59) peut s'écrire

$$\sum_{i=0}^{n} \psi_i(t)\, f_i(\hat{x}(t), u(t)) \leqslant \sum_{i=0}^{n} \psi_i(t)\, f_i(\hat{x}(t), \hat{u}(t)). \qquad (17.60)$$

Dans ce qui précède (10) nous avons désigné par $\widetilde{H}$ la fonction des variables $x_1, \ldots, x_n, \psi_0, \psi_1, \ldots, \psi_n, u_1, \ldots, u_r$:

$$\widetilde{H}(\psi, x, u) = \sum_{i=0}^{n} \psi_i f_i(x, u).$$

C'est pourquoi l'inégalité (60) devient

$$\widetilde{H}(\psi, \hat{x}, u) \leqslant \widetilde{H}(\psi, \hat{x}, \hat{u}). \qquad (17.61)$$

Remarquons encore que puisque d'après (30) et (36),

$$\frac{\partial f_i(x, u)}{\partial x_0} = \frac{\partial f_i(x_1, x_2, \ldots, x_n, u_1, \ldots, u_r)}{\partial x_0} = 0 \qquad (i = 0, 1, \ldots, n), \qquad (17.62)$$

(51) entraîne que

$$\frac{d\psi_0}{dt} = 0, \qquad (17.63)$$

d'où on trouve en vertu de (55)

$$\psi_0(t) = -1 \qquad (t_0 \leqslant t \leqslant T). \qquad (17.64)$$

Nous avons vu ainsi que si la commande $\hat{u}(t)$ est optimale, on a les conditions (61) et (64) qui sont des conditions nécessaires d'optimalité; ces dernières traduisent le contenu du théorème sur la condition nécessaire d'optimalité (principe du maximum) du problème à temps fixé et à extrémité libre de la trajectoire.

La démonstration du théorème 1 pour le problème général de la commande optimale est donnée dans la monographie [72].

§ 18. Principe du maximum pour des systèmes non autonomes

1. Théorème de la condition nécessaire d'optimalité. Considérons le système décrit par les équations différentielles

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r, t) = f_j(x, u, t) \qquad (18.1)$$

$$(j = 1, \ldots, n)$$

auxquelles est équivalente l'équation vectorielle

$$\frac{dx}{dt} = f(x, u, t), \qquad (18.2)$$

où x, u et f sont les vecteurs

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u, t) = \begin{bmatrix} f_1(x, u, t) \\ \cdots\cdots \\ f_n(x, u, t) \end{bmatrix}. \qquad (18.3)$$

On demande d'amener le système du point $x(t_0)$ de l'espace de phase X en un point donné x^*. L'instant t_1 où le point représentatif rejoint x^* n'est pas fixé à l'avance.

La commande $u = u(t)$ doit vérifier les contraintes

$$u \in \Omega; \qquad (18.4)$$

de plus, le domaine Ω est supposé indépendant du temps.

En tenant compte des contraintes (4), il faut choisir une commande u telle qu'elle minimise la fonctionnelle

$$Q = \int_{t_0}^{t_1} f_0(x(t), u(t), t)\, dt. \qquad (18.5)$$

La commande u vérifiant ces conditions, la trajectoire qui lui est associée et l'intervalle de temps $t_1 - t_0$ sont considérés comme *optimaux*.

Comme dans ce qui précède, désignons par $x_0(t)$ la fonction scalaire définie par l'équation différentielle

$$\frac{dx_0}{dt} = f_0(x_1, \ldots, x_n, u_1, \ldots, u_r, t) = f_0(x, u, t) \qquad (18.6)$$

et par la condition initiale

$$x_0(t_0) = 0. \tag{18.7}$$

D'après (6) et (7), la fonctionnelle (5) peut s'écrire

$$Q = x_0(t_1). \tag{18.8}$$

Introduisons encore la fonction scalaire x_{n+1} définie par l'équation différentielle

$$\frac{dx_{n+1}}{dt} = 1 \tag{18.9}$$

et par la condition initiale

$$x_{n+1}(t_0) = t_0. \tag{18.10}$$

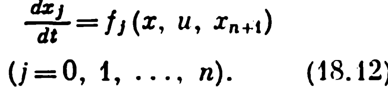
Il est clair que

$$x_{n+1} = t. \tag{18.11}$$

L'espace des variables $x_1, x_2, \ldots, x_n, x_{n+1}$ est désigné par X^*.

Fig. 18.1

Le système d'équations (1) et l'équation (6) peuvent s'écrire en commun



$$\frac{dx_j}{dt} = f_j(x, u, x_{n+1})$$
$$(j = 0, 1, \ldots, n). \tag{18.12}$$

Voici l'énoncé de notre problème. Chercher la trajectoire optimale qui relie dans l'espace X^* le point $(x_0(t_0), x_2(t_0), \ldots, x_n(t_0), t_0)$ avec un certain point de la droite S_1 (fig. 18.1) qui passe par le point $(x_1^*, x_2^*, \ldots, x_n^*, 0)$ parallèlement à l'axe x_{n+1}.

Le problème concerné est ainsi ramené à un problème optimal autonome à extrémité gauche fixée mais à extrémité droite libre. L'extrémité droite de la trajectoire étant libre, l'application du théorème 1 (§ 17) ne peut être immédiate.

Le système d'équations auxiliaire (17.9) est ici de la forme

$$\frac{d\psi_k}{dt} = -\sum_{i=0}^{n} \frac{\partial f_i(x, u, t)}{\partial x_k} \psi_i \qquad (k = 0, 1, \ldots, n), \tag{18.13}$$

$$\frac{d\psi_{n+1}}{dt} = -\sum_{i=0}^{n} \frac{\partial f_i(x, u, t)}{\partial t} \psi_i. \tag{18.14}$$

Introduisons la notation analogue à (17.10)

$$\tilde{H}(\psi, x, t, u) = \sum_{i=0}^{n} \psi_i f_i(x, u, t). \tag{18.15}$$

Les équations (12) et (13) peuvent se mettre sous la forme

$$\frac{dx_i}{dt} = \frac{\partial \widetilde{H}}{\partial \psi_i}, \quad \frac{d\psi_i}{dt} = -\frac{\partial \widetilde{H}}{\partial x_i} \qquad (i = 0, 1, \ldots, n). \qquad (18.16)$$

La condition nécessaire d'optimalité d'un système non autonome est donnée par le théorème qui suit.

Théorème (principe du maximum pour les systèmes non autonomes [72]). *Soit* $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ *une commande admissible telle que la trajectoire qui lui est associée* $x(t)$ *du système* (1), *issue à l'instant* t_0 *du point* $x(t_0)$, *passe à l'instant* t_1 *par le point* x^*. *Pour que la commande* $u(t)$ *et la trajectoire* $x(t)$ *soient optimales, il faut qu'il existe un vecteur fonction continu non nul* $\psi(t) = (\psi_0(t), \psi_1(t), \ldots, \psi_n(t))$, *associé aux fonctions* $u(t)$ *et* $x(t)$ *tel que*

1°. *Pour tout instant* t $(t_0 \leqslant t \leqslant t_1)$, *qui est un point de continuité de la commande* $u(t)$, *la fonction* $\widetilde{H}(\psi(t), x(t), t, u)$ *de la variable* $u \in \Omega$ *atteint au point* $u = u(t)$ *le maximum*

$$\widetilde{H}(\psi(t), x(t), t, u(t)) = M(\psi(t), x(t), t). \qquad (18.17)$$

2°. *Les relations*

$$\psi_0(t) = \text{const} \leqslant 0, \qquad (18.18)$$

$$M(\psi(t), x(t), t) = \int_{t_1}^{t} \sum_{i=1}^{n} \frac{\partial f_i(x(t), u(t), t)}{\partial t} \psi_i(t)\, dt \qquad (18.19)$$

soient remplies.

Il s'avère ensuite que si les fonctions $\psi(t)$, $x(t)$, $u(t)$ *vérifient le système* (12), (13) *et la condition* 1°, *la fonction* $\psi_0(t)$ *de la variable* t *est constante, et la fonction* $M(\psi(t), x(t), t)$ *ne peut différer de l'intégrale de* (19) *que d'une constante*; *il suffit donc que les relations* (18) *et* (19) *soient vérifiées à un instant quelconque* t $(t_0 \leqslant t \leqslant t_1)$; *par exemple au lieu de* (18) *et* (19), *il suffit de vérifier les relations*

$$\psi_0(t_1) \leqslant 0, \quad M(\psi(t_1), x(t_1), t_1) = 0. \qquad (18.20)$$

La comparaison du théorème qui vient d'être énoncé avec le théorème 1 (§ 17) montre que pour les systèmes non autonomes décrits par les équations du type (1), le principe du maximum est également vrai.

La différence ne consiste que dans le fait que pour les systèmes non autonomes la fonction

$$M(\psi(t), x(t), t) = \widetilde{H}(\psi(t), x(t), t, u(t))$$

n'est plus constante, étant définie par l'expression (19).

2. Démonstration du théorème de la condition nécessaire d'optimalité d'un système non autonome à entrée linéaire de la commande. Pour le cas particulier d'une entrée linéaire de la commande on peut donner une démonstration immédiate du théorème de la condition nécessaire d'optimalité d'un système non autonome [27, 72].

Examinons le système régi par les équations différentielles scalaires

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, t) + \sum_{l=1}^{r} B_{jl}(t)\, u_l \qquad (j = 1, \ldots, n) \tag{18.21}$$

auxquelles est équivalente l'équation différentielle vectorielle

$$\frac{dx}{dt} = f(x, t) + B(t)\, u, \tag{18.22}$$

où

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad f(x, t) = \begin{bmatrix} f_1(x, t) \\ \cdots\cdots \\ f_n(x, t) \end{bmatrix}, \qquad B(t) = \begin{bmatrix} B_{11}(t) & \ldots & B_{1r}(t) \\ \ldots & \ldots & \ldots \\ B_{n1}(t) & \ldots & B_{nr}(t) \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}. \tag{18.23}$$

Il faut faire passer le système du point $x(t_0)$ au point donné x^*. L'instant t_1 où le point représentatif tombe en x^* n'est pas fixé à l'avance.

Les commandes $u_l(t)$ doivent vérifier les contraintes

$$|u_l(t)| \leqslant m_l, \qquad (l = 1, \ldots, r), \tag{18.24}$$

et il faut les choisir telles que le système soit amené en x^* en minimisant la fonctionnelle

$$Q = \int_{t_0}^{t_1} \left[f_0(x(t), t) + \sum_{l=1}^{r} B_{0l}(t)\, u_l(t) \right] dt. \tag{18.25}$$

Désignons par $x_0(t)$ la fonction définie par l'équation différentielle

$$\frac{dx_0}{dt} = f_0(x_1, \ldots, x_n, t) + \sum_{l=1}^{r} B_{0l}(t)\, u_l(t) \tag{18.26}$$

et la condition initiale

$$x_0(t_0) = 0. \tag{18.27}$$

La fonctionnelle Q se met alors sous la forme

$$Q = x_0(t_1). \tag{18.28}$$

Les variables auxiliaires $\psi_0, \psi_1, \ldots, \psi_n$ vérifient le système d'équations différentielles

$$\frac{d\psi_k}{dt} = -\sum_{i=0}^{n} \frac{\partial f_i(x_1, \ldots, x_n, t)}{\partial x_k} \psi_i \qquad (k=0, 1, \ldots, n). \tag{18.29}$$

Dans le problème considéré la fonction $\widetilde{H}$ s'écrit

$$\widetilde{H} = \sum_{i=0}^{n} \psi_i \left[f_i(x_1, \ldots, x_n, t) + \sum_{l=1}^{r} B_{il}(t) u_l \right]$$

ou

$$\widetilde{H} = \sum_{i=0}^{n} \psi_i f_i(x_1, \ldots, x_n, t) + \sum_{l=1}^{r} u_l \left[\sum_{i=0}^{n} B_{il}(t) \psi_i \right]. \tag{18.30}$$

Désignons par $\hat{u}(t)$ la commande optimale, par $\hat{x}(t)$ et $\hat{\psi}(t)$ les vecteurs des coordonnées de phase et des variables auxiliaires qui lui sont associés.

Montrons que la commande optimale aura la forme suivante qui correspond au théorème du principe du maximum

$$\hat{u}_l(t) = m_l \operatorname{sign} \sum_{i=0}^{n} B_{il}(t) \hat{\psi}_i(t) \qquad (l=1, \ldots, r) \tag{18.31}$$

ou

$$\hat{u}_l(t) = m_l \operatorname{sign} R_l(t) \qquad (l=1, \ldots, r), \tag{18.32}$$

avec

$$R(t) = \widetilde{B}^*(t) \hat{\psi}(t), \tag{18.33}$$

$$\widetilde{B}(t) = \begin{bmatrix} B_{01}(t) & \ldots & B_{0r}(t) \\ B_{11}(t) & \ldots & B_{1r}(t) \\ \ldots & \ldots & \ldots \\ B_{n1}(t) & \ldots & B_{nr}(t) \end{bmatrix}, \quad \hat{\psi}(t) = \begin{bmatrix} \hat{\psi}_0(t) \\ \hat{\psi}_1(t) \\ \ldots \\ \hat{\psi}_n(t) \end{bmatrix}. \tag{18.34}$$

Donnons à la commande optimale $\hat{u}_l(t)$ l'accroissement $\delta u_l(t)$ qui vérifie la condition

$$|\hat{u}_l(t) + \delta u_l(t)| < m_l \qquad (l = 1, \ldots, r). \tag{18.35}$$

Appelons la fonction $\delta u_l(t)$ *variation admissible* de la commande optimale.

A la commande $\hat{u}_l(t) + \delta u_l(t)$ $(l = 1, \ldots, r)$ correspond la solution $\hat{x}_k(t) + \delta x_k(t)$ $(k = 0, 1, \ldots, n)$ des équations différentielles (21) et (26), où $\hat{x}_k(t)$ $(k = 0, 1, \ldots, n)$ est la solution de ces équations pour $u_l(t) = \hat{u}_l(t)$ $(l = 1, \ldots, r)$.

L'état initial du système $x_0(t_0)$, $x_1(t_0)$, ..., $x_n(t_0)$ étant fixé, il vient

$$\delta x_k(t_0) = 0 \qquad (k = 0, 1, \ldots, n). \tag{18.36}$$

Ainsi, suivant (21) et (26), on a le système d'équations différentielles

$$\frac{d}{dt}(\hat{x}_j + \delta x_j) = f_j(x_1 + \delta x_1, \ldots, \hat{x}_n + \delta x_n, t) + \\ + \sum_{l=1}^{r} B_{jl}(t)(\hat{u}_l + \delta u_l) \qquad (j = 0, 1, \ldots, n). \tag{18.37}$$

Comme

$$f_j(\hat{x}_1 + \delta x_1, \ldots, \hat{x}_n + \delta x_n, t) = \\ = f_j(\hat{x}_1, \ldots, \hat{x}_n, t) + \sum_{k=0}^{n} \frac{\partial f_j(\hat{x}_1, \ldots, \hat{x}_n, t)}{\partial x_k} \delta x_k + \\ + o(\delta x_0, \delta x_1, \ldots, \delta x_n) \qquad (j = 0, 1, \ldots, n), \tag{18.38}$$

les équations (37) deviennent

$$\frac{d\hat{x}_j}{dt} + \frac{d}{dt}\delta x_j = f_j(\hat{x}_1, \ldots, \hat{x}_n, t) + \sum_{l=1}^{r} B_{jl}(t)\hat{u}_l + \\ + \sum_{k=0}^{n} \frac{\partial f_j(\hat{x}_1, \ldots, \hat{x}_n, t)}{\partial x_k} \delta x_k + \sum_{l=1}^{r} B_{jl}(t)\delta u_l + o(\delta x_0, \delta x_1, \ldots, \delta x_n) \\ (j = 0, 1, \ldots, n). \tag{18.39}$$

Comme il résulte des relations (21) et (26), le premier terme du premier membre de l'équation (39) se réduit avec les deux premiers termes du second membre de ces équations. Dans le cas de petites variations admissibles de la commande $\delta u_l(t)$, les variations $\delta x_l(t)$ $(l = 0, 1, \ldots, n)$ seront également petites. Alors, en négligeant dans les équations (39) l'ensemble des termes du deuxième degré et des degrés plus élevés par rapport à $\delta x_0, \delta x_1, \ldots, \delta x_n$, on obtient le système d'équations différentielles linéaires

$$\frac{d}{dt}\delta x_j = \sum_{k=0}^{n} \frac{\partial f_j(\hat{x}_1, \ldots, \hat{x}_n, t)}{\partial x_k} \delta x_k + \sum_{l=1}^{r} B_{jl}(t)\delta u_l(t) \tag{18.40}$$

$$(j = 0, 1, \ldots, n).$$

Les équations (40) représentent un système d'équations aux variations du problème considéré. Puisque $\hat{x}_1(t), \ldots, \hat{x}_n(t)$ déterminent la trajectoire optimale, les fonctions $\frac{\partial f_j(\hat{x}_1, \ldots, \hat{x}_n, t)}{\partial x_k}$ sont certaines fonctions du temps.

Désignons par $A\ (t)$ la matrice $(n+1)\times(n+1)$ dont les éléments sont les fonctions $\frac{\partial f_j(\hat{x}_1, \ldots, \hat{x}_n, t)}{\partial x_k}$:

$$A\,(t)=\begin{bmatrix} \frac{\partial f_0}{\partial x_0} & \frac{\partial f_0}{\partial x_1} & \frac{\partial f_0}{\partial x_2} & \cdots & \frac{\partial f_0}{\partial x_n} \\ \frac{\partial f_1}{\partial x_0} & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_0} & \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}. \qquad (18.41)$$

Introduisons la notation ξ pour le vecteur de dimension $(n+1)$

$$\xi=\begin{bmatrix} \xi_0 \\ \xi_1 \\ \cdots \\ \xi_n \end{bmatrix}, \qquad (18.42)$$

où

$$\xi_j=\delta x_j \qquad (j=0,\ 1,\ \ldots,\ n). \qquad (18.43)$$

Le système d'équations aux variations peut être mis sous la forme de l'équation différentielle vectorielle

$$\dot{\xi}=A\ (t)\ \xi+\widetilde{B}\ (t)\ \delta u\ (t), \qquad \xi\ (t_0)=0, \qquad (18.44)$$

où la matrice $\widetilde{B}\ (t)$ est de la forme (34).

La matrice $A^*\ (t)$ construite par transposition de la matrice (41) s'écrit:

$$A^*\ (t)=\begin{bmatrix} \frac{\partial f_0}{\partial x_0} & \frac{\partial f_1}{\partial x_0} & \frac{\partial f_2}{\partial x_0} & \cdots & \frac{\partial f_n}{\partial x_0} \\ \frac{\partial f_0}{\partial x_1} & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_0}{\partial x_n} & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}. \qquad (18.45)$$

D'après (29), les variables auxiliaires $\hat{\psi}_0, \hat{\psi}_1, \ldots, \hat{\psi}_n$ associées à la commande optimale $\hat{u}_l\ (t)$ vérifient le système d'équations différentielles

$$\frac{d\hat{\psi}_k}{dt}=-\sum_{i=0}^{n}\frac{\partial f_i\,(\hat{x}_1, \ldots, \hat{x}_n, t)}{\partial x_k}\,\hat{\psi}_i \qquad (k=0,\ 1,\ \ldots,\ n). \qquad (18.46)$$

Ce système peut être remplacé par l'équation différentielle vectorielle

$$\frac{d}{dt}\,\hat{\psi}=-A^*\,(t)\,\hat{\psi}. \qquad (18.47)$$

Désignons par $\Xi(t)$ la matrice fondamentale des solutions de l'équation différentielle vectorielle (47) vérifiant la condition $\Xi(t_0) = E$, où E est la matrice unité.

Par $\Lambda(t)$ désignons la matrice fondamentale des solutions de l'équation différentielle vectorielle

$$\dot{\zeta} = A(t)\,\zeta, \tag{18.48}$$

qui satisfait à la condition $\Lambda(t_0) = E$.

Comme d'après (47) et (48),

$$\begin{aligned}\frac{d}{dt}\langle\hat{\psi}(t),\ \zeta(t)\rangle &= \left\langle\frac{d}{dt}\hat{\psi}(t),\ \zeta(t)\right\rangle + \left\langle\hat{\psi}(t),\ \frac{d}{dt}\zeta(t)\right\rangle = \\ &= \langle -A^*(t)\hat{\psi}(t),\ \zeta(t)\rangle + \langle\hat{\psi}(t),\ A(t)\zeta(t)\rangle = \\ &= -\hat{\psi}^*(t)A(t)\zeta(t) + \hat{\psi}^*(t)A(t)\zeta(t) = 0,\end{aligned} \tag{18.49}$$

il vient

$$\langle\hat{\psi}(t),\ \zeta(t)\rangle = \text{const}, \tag{18.50}$$

d'où l'on déduit que

$$\langle\hat{\psi}(t),\ \zeta(t)\rangle = \langle\hat{\psi}(t_0),\ \zeta(t_0)\rangle. \tag{18.51}$$

En remplaçant $\hat{\psi}(t)$ et $\zeta(t)$ par les solutions des équations (47) et (48)

$$\hat{\psi}(t) = \Xi(t)\,\hat{\psi}(t_0), \quad \zeta(t) = \Lambda(t)\,\zeta(t_0),$$

on réduit la relation (51) à la forme

$$[\Xi(t)\,\hat{\psi}(t_0)]^*\,\Lambda(t)\,\zeta(t_0) = \hat{\psi}^*(t_0)\,\zeta(t_0)$$

ou

$$\hat{\psi}^*(t_0)\,\Xi^*(t)\,\Lambda(t)\,\zeta(t_0) = \hat{\psi}^*(t_0)\,\zeta(t_0). \tag{18.52}$$

La relation (52) entraîne

$$\Xi^*(t)\,\Lambda(t) \equiv E, \quad \Xi^*(t) = \Lambda^{-1}(t). \tag{18.53}$$

En transposant les matrices du premier et du second membre de (53), on obtient

$$\Lambda^*(t)\,\Xi(t) \equiv E, \quad \Lambda^*(t) = \Xi^{-1}(t). \tag{18.54}$$

D'une façon analogue à (49), on trouve des équations (44) et (47) que

$$\frac{d}{dt}\langle\xi(t),\ \hat{\psi}(t)\rangle = \langle\widetilde{B}^*(t)\,\hat{\psi}(t),\ \delta u(t)\rangle, \tag{18.55}$$

d'où, en intégrant, on obtient

$$\langle\xi(t_1),\ \hat{\psi}(t_1)\rangle - \langle\xi(t_0),\ \hat{\psi}(t_0)\rangle = \int_{t_0}^{t_1}\langle\widetilde{B}^*(\tau)\,\hat{\psi}(\tau),\ \delta u(\tau)\rangle\,d\tau. \tag{18.56}$$

Puisque d'après (44) $\xi(t_0) = 0$, la relation (56) devient

$$\langle \xi(t_1),\ \hat{\psi}(t_1)\rangle = \int_{t_0}^{t_1} \langle \widetilde{B}^*(\tau)\,\hat{\psi}(\tau),\ \delta u(\tau)\rangle\, d\tau. \tag{18.57}$$

Montrons que si la condition (31) est respectée, on a la relation

$$\langle \xi(t_1),\ \hat{\psi}(t_1)\rangle \leqslant 0, \tag{18.58}$$

où $\xi(t_1)$ est un vecteur quelconque de l'ensemble des vecteurs associés à l'ensemble des variations admissibles $\delta u(t)$.

En effet, conformément à (31), pour

$$\sum_{i=0}^{n} B_{i\mu}(t)\,\hat{\psi}_i(t) > 0 \qquad (t \in [t^*,\ t^{**}] \subset [t_0,\ t_1])$$

on a $\hat{u}_\mu(t) = m_\mu$ $(t \in [t^*,\ t^{**}])$ et la variation admissible $\delta u_\mu(t)$ satisfait à la condition

$$-2m_\mu < \delta u_\mu(t) \leqslant 0.$$

Pour

$$\sum_{i=0}^{n} B_{i\sigma}(t)\,\hat{\psi}_i(t) < 0 \qquad (t \in [t_*,\ t_{**}] \subset [t_0,\ t_1])$$

on a, d'après (31), $\hat{u}_\sigma(t) = -m_\sigma$ $(t \in [t_*,\ t_{**}])$ et la variation admissible satisfait à la condition

$$0 \leqslant \delta u_\sigma(t) < 2m_\sigma.$$

De plus

$$\int_{t_0}^{t_1} \langle \widetilde{B}^*(\tau)\,\hat{\psi}(\tau),\ \delta u(\tau)\rangle\, d\tau \leqslant 0,$$

d'où l'on déduit d'après (57) la relation (58).

On voit également sans peine que si la condition (31) n'est pas observée, il existe des variations admissibles telles que

$$\langle \xi(t_1)\ \hat{\psi}(t_1)\rangle > 0.$$

En effet, supposons qu'en dehors de l'intervalle de temps $[t^*,\ t^{**}] \subset$ $\subset [t_0,\ t_1]$, la variation de la commande $\delta u(t) \equiv 0$. Dans l'intervalle de temps $[t^*,\ t^{**}]$

$$\delta u_\mu(t) \neq 0,\ \ \delta u_j(t) \equiv 0 \qquad (j = 1,\ \ldots,\ \mu - 1,\ \mu + 1,\ \ldots,\ r).$$

Soient

$$\sum_{i=0}^{n} B_{i\mu}(t)\,\hat{\psi}_i(t) > 0,\ \ u_\mu(t) < m_\mu \qquad (t \in [t^*,\ t^{**}]),$$

c'est-à-dire la condition 31) n'est pas remplie.

Choisissons comme variation admissible de la commande

$$\delta u_\mu(t) = \nu > 0 \qquad (t \in [t^*, t^{**}]),$$

où ν vérifie la condition

$$u_\mu(t) + \nu < m_\mu \qquad (t \in [t^*, t^{**}]).$$

L'expression (57) se met alors sous la forme

$$\langle \xi(t_1), \hat{\psi}(t_1) \rangle = \int_{t^*}^{t^{**}} \Big[\sum_{i=0}^{n} B_{i\mu}(\tau) \hat{\psi}_i(\tau) \Big] \nu \, d\tau > 0.$$

Ainsi la relation (58) n'a lieu que si la condition (31) est observée.

Les variations admissibles $\delta u(t)$ de la commande optimale doivent vérifier la relation (35), dont il résulte que les vecteurs $\delta u(t)$ sont bornés en norme. Puisque nous partons du système linéaire d'équations aux variations (44), nous nous bornerons aux variations $\delta u(t)$ suffisamment petites vérifiant aussi bien la condition (35) que la condition

$$|\delta u_l(t)| \leqslant \varepsilon \qquad (l = 1, \ldots, r), \tag{18.59}$$

où ε est une grandeur suffisamment petite.

A chaque variation admissible $\delta u(t)$ correspond une solution $\xi(t)$ de l'équation différentielle vectorielle (44). A l'instant t_1, cette solution prend la valeur $\xi(t_1)$. Dans l'espace $(x_0, x_1, \ldots, x_n)$ de dimension $(n+1)$ menons à partir du point $\tilde{x}(t_1) = (x_0(t_1), \hat{x}(t_1))$ le vecteur $\xi(t_1)$. Appelons l'ensemble des extrémités des vecteurs $\xi(t_1)$ associés aux variations admissibles $\delta u(t)$ *ensemble d'atteignabilité* et désignons-le par $K(t_1)$. Puisque d'après (59) les variations admissibles $\delta u(t)$ sont bornées, les vecteurs $\xi(t_1)$ le sont également et, par conséquent, l'ensemble $K(t_1)$ est borné.

On peut montrer que $K(t_1)$ est un ensemble convexe. A cet effet il faut que tous les points

$$(1-\lambda)\xi^*(t_1) + \lambda\xi^{**}(t_1) \qquad (0 \leqslant \lambda \leqslant 1)$$

du segment de droite reliant les deux points $\xi^*(t_1)$ et $\xi^{**}(t_1)$ de l'ensemble $K(t_1)$ appartiennent à cet ensemble. Les variations de la commande $\delta u^*(t)$ et $\delta u^{**}(t)$ auxquelles correspondent $\xi^*(t_1)$ et $\xi^{**}(t_1)$ vérifient les conditions (35) et (59). Définissons $\delta u_\lambda(t)$ $(t \in [t_0, t_1])$ par la relation

$$\delta u_\lambda(t) = (1-\lambda)\,\delta u^*(t) + \lambda\delta u^{**}(t) \qquad (0 \leqslant \lambda \leqslant 1).$$

La variation de la commande $\delta u_\lambda(t)$ satisfait aux conditions (35) et (59), c'est-à-dire le point $\xi_\lambda(t_1) \in K(t_1)$. La solution $\xi_\lambda(t)$ qui correspond à la variation $\delta u_\lambda(t)$ et à la condition initiale $\xi_\lambda(t_0) = 0$ est, en vertu de (44),

$$\xi_\lambda(t) = \Lambda(t) \int_{t_0}^{t} \Lambda^{-1}(\tau) \tilde{B}(\tau) \delta u_\lambda(\tau) \, d\tau =$$

$$= (1-\lambda)\Lambda(t)\int_{t_0}^{t}\Lambda^{-1}(\tau)\widetilde{B}(\tau)\,\delta u^{*}(\tau)\,d\tau +$$

$$+\lambda\Lambda(t)\int_{t_0}^{t}\Lambda^{-1}(\tau)\widetilde{B}(\tau)\,\delta u^{**}(\tau)\,d\tau = (1-\lambda)\,\xi^{*}(t)+\lambda\xi^{**}(t).$$

On en déduit que

$$\xi_\lambda(t_1) = (1-\lambda)\,\xi^{*}(t_1)+\lambda\xi^{**}(t_1);$$

$K(t_1)$ est donc un ensemble convexe borné.

Montrons encore que le point $\widetilde{x}(t_1) = (\hat{x}_0(t_1), \hat{x}(t_1))$, associé à la variation $\delta u(t) \equiv 0$ ($t \in [t_0, t_1]$), est situé sur la frontière de l'ensemble $K(t_1)$. En effet, le point $\widetilde{x}(t_1)$ de l'espace $(x_0, x_1, \ldots, x_n)$ repose sur la droite L qui passe par le point $(0, \hat{x}_1(t_1), \ldots, \hat{x}_n(t_1))$ parallèlement à l'axe x_0. Si le point $(\hat{x}_0(t_1), \hat{x}_1(t_1), \ldots, x_n(t_1))$ était un point intérieur de l'ensemble convexe borné $K(t_1)$, il existerait un segment L^* de la droite L d'extrémité $(\hat{x}_0(t_1), \hat{x}_1(t_1), \ldots$ $\ldots, \hat{x}_n(t_1))$, dont tous les points appartiendraient à l'ensemble $K(t_1)$ et auraient la coordonnée x_0, plus petite que la coordonnée $\hat{x}_0(t_1)$ du point frontière de la trajectoire optimale. Alors, une sphère de rayon ρ assez petit et de centre $(\hat{x}_0(t_1), \hat{x}_1(t_1), \ldots, \hat{x}_n(t_1))$, dont tous les points appartiennent à l'ensemble $K(t_1)$ couperait le segment L^* au point $(x_0^*, \hat{x}_1(t_1), \ldots, \hat{x}_n(t_1))$, où $x_0^* < \hat{x}_0(t_1)$, ce qui contredirait l'optimalité de la trajectoir e$(\hat{x}_0(t), \hat{x}_1(t), \ldots, \hat{x}_n(t))$ $(t \in [t_0, t_1])$.

Il s'ensuit que le point $\widetilde{x}(t_1) = (\hat{x}_0(t_1), \hat{x}(t_1))$ appartient à la frontière de l'ensemble convexe borné $K(t_1)$.

On sait [15] que par tout point frontière d'un ensemble convexe no deut mener au moins un hyperplan tel que tous les points de cet ensemble se situeront d'un seul côté de cet hyperplan. Cet hyperplan est dit *hyperplan d'appui*. Par un point frontière anguleux on peut mener plus d'un hyperplan d'appui. Désignons par M l'hyperplan d'appui de l'ensemble $K(t_1)$ passant par le point frontière $\widetilde{x}(t_1) = (\hat{x}_0(t_1), \hat{x}(t_1))$. Désignons par N la normale unité à l'hyperplan M en $\widetilde{x}(t_1) = (\hat{x}_0(t_1), \hat{x}(t_1))$, dirigée dans le demi-espace ne contenant pas d'ensemble $K(t_1)$. De ce qui vient d'être dit il en résulte que pour tout vecteur $\xi(t_1) \in K(t_1)$ on a la relation

$$\langle \xi(t_1), N\rangle \leqslant 0. \qquad (18.60)$$

Pour achever la démonstration du théorème, il faut démontrer l'existence d'un vecteur fonction $\hat{\psi}(t)$ vérifiant la condition (17).

Posons

$$\hat{\psi}(t_1) = N. \tag{18.61}$$

Alors, comme nous l'avons montré plus haut (relation (58)), pour remplir la condition (60) il faut que la commande $\hat{u}(t)$ soit définie par la formule (31), où $\hat{\psi}(t)$ est la solution du système d'équations différentielles (46) vérifiant la condition aux limites (61).

Remarquons à titre d'exemple que pour un problème à temps fixé T et à extrémité libre de la trajectoire le plan d'appui M de l'ensemble d'atteignabilité $K(T)$ est perpendiculaire à l'axe x_0. C'est pourquoi la normale N au plan M, dirigée dans le demi-espace ne contenant pas d'ensemble $K(T)$, est $N = (-1, 0, \ldots, 0)$ et, par suite, d'après (61), dans ce problème $\psi(T) = (-1, 0, \ldots, 0)$, comme l'indique la formule (17.55).

Les raisonnements exposés ici à l'aide des équations linéaires aux variations (44) sont vrais si les *variations* $\delta u(t)$ *sont assez petites*, c'est-à-dire si l'on n'étudie que les trajectoires assez proches de la trajectoire optimale.

Ainsi nous avons démontré que la condition (31) est une condition *nécessaire* d'optimalité.

La question des conditions suffisantes d'optimalité sous la forme du principe du maximum impose une étude spéciale.

3. Systèmes linéaires non autonomes. Réduction du problème de la réponse en temps minimal à un problème aux limites. Introduisons au préalable quelques notations. Considérons deux systèmes décrits par les équations différentielles vectorielles

$$\frac{d\xi}{dt} = A(t)\,\xi, \tag{18.62}$$

$$\frac{d\eta}{dt} = -A^*(t)\,\eta, \tag{18.63}$$

où la matrice $A^*(t)$ est la transposée de $A(t)$. Désignons par $\theta(t)$ la matrice fondamentale des solutions de l'équation (62), et par $X(t)$ la matrice

$$X(t) = \theta(t)\,\theta^{-1}(0). \tag{18.64}$$

La solution du problème de Cauchy pour l'équation différentielle vectorielle (62) est

$$\xi(t) = X(t)\,\xi(0). \tag{18.65}$$

La solution du problème de Cauchy pour l'équation vectorielle (63) s'écrit

$$\eta(t) = \Xi(t)\,\eta(0). \tag{18.66}$$

Ici $\Xi(t)$ est une matrice carrée qui satisfait à la condition

$$\Xi(0) = E, \tag{18.67}$$

où E est la matrice unité.

Comme nous l'avons montré précédemment (53), les matrices $X(t)$ et $\Xi(t)$ vérifient la condition

$$\Xi^*(t) = X^{-1}(t). \qquad (18.68)$$

Considérons maintenant le système de commande décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = A(t)\,x + B(t)\,u, \qquad (18.69)$$

où x, $A(t)$, $B(t)$, u sont les matrices

$$\left.\begin{aligned} x &= \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \\ A(t) &= \begin{bmatrix} A_{11}(t) & \dots & A_{1n}(t) \\ \dots & \dots & \dots \\ A_{n1}(t) & \dots & A_{nn}(t) \end{bmatrix}, \\ B(t) &= \begin{bmatrix} B_{11}(t) & \dots & B_{1r}(t) \\ \dots & \dots & \dots \\ B_{n1}(t) & \dots & B_{nr}(t) \end{bmatrix}. \end{aligned}\right\} \qquad (18.70)$$

L'équation vectorielle (69) est équivalente au système d'équations scalaires

$$\frac{dx_j}{dt} = \sum_{k=1}^{n} A_{jk}(t)\,x_k + \sum_{l=1}^{r} B_{jl}(t)\,u_l \qquad (j = 1, \dots, n). \qquad (18.71)$$

Les commandes u_l sont soumises aux contraintes

$$-1 \leqslant u_l \leqslant 1 \qquad (l = 1, \dots, r). \qquad (18.72)$$

On demande de faire passer ce système en un temps minimal du point $x(0)$, où le système se trouve à l'instant $t_0 = 0$, au point $x = 0$.

Désignons par ψ le vecteur de dimension n

$$\psi = \begin{bmatrix} \psi_1 \\ \cdots \\ \psi_n \end{bmatrix} \qquad (18.73)$$

et introduisons d'après (15) la fonction

$$H(\psi, x, u) = \sum_{j=1}^{n} \sum_{k=1}^{n} \psi_j A_{jk}(t)\,x_k + \sum_{l=1}^{r} u_l \Big(\sum_{j=1}^{n} \psi_j B_{jl}(t)\Big). \qquad (18.74)$$

En vertu de (13), les fonctions ψ_j $(j = 1, \dots, n)$ vérifient les équations différentielles

$$\frac{d\psi_j}{dt} = -\sum_{k=1}^{n} A_{kj}(t)\,\psi_k \qquad (j = 1, \dots, n), \qquad (18.75)$$

équivalentes à l'équation différentielle vectorielle

$$\frac{d\psi}{dt} = -\dot{A}^* (t) \psi. \tag{18.76}$$

Conformément au théorème du principe du maximum, la commande optimale a pour expression

$$u_l (t) = \operatorname{sign} \sum_{j=1}^{n} B_{jl} (t) \psi_j (t) \qquad (l = 1, \ldots, r). \tag{18.77}$$

Les relations (77) peuvent s'écrire

$$u_l (t) = \operatorname{sign} R_l (t) \qquad (l = 1, \ldots, r), \tag{18.78}$$

où par R (t) on désigne le vecteur de dimension r qui est le produit des matrices

$$R (t) = B^* (t) \psi (t). \tag{18.79}$$

Introduisons la notation Y (t) pour la matrice

$$Y (t) = X^{-1} (t) B (t). \tag{18.80}$$

Comme en vertu de (66) et (63), la solution de l'équation vectorielle (76) est

$$\psi (t) = \Xi (t) \psi (0), \tag{18.81}$$

l'expression (79) peut s'écrire, compte tenu de (68),

$$R (t) = B^* (t) \Xi (t) \psi (0) = [\Xi^* (t) B (t)]^* \psi (0) = \\ = [X^{-1} (t) B (t)]^* \psi (0) = Y^* (t) \psi (0). \tag{18.82}$$

D'après (82), les éléments du vecteur R (t) sont

$$R_l (t) = \sum_{j=1}^{n} Y_{jl} (t) \psi_j (0) \qquad (l = 1, \ldots, r). \tag{18.83}$$

De cette façon d'après (78) et (83), la commande optimale est

$$u_l (t) = \operatorname{sign} \sum_{j=1}^{n} Y_{jl} (t) \psi_j (0) \qquad (l = 1, \ldots, r). \tag{18.84}$$

Ecrivons maintenant la solution du problème de Cauchy pour l'équation vectorielle (69)

$$x (t) = X (t) x (0) + X (t) \int_0^t X^{-1} (\tau) B (\tau) u (\tau) d\tau. \tag{18.85}$$

Désignons par t_1 l'instant où le système (69) est recalé à l'origine

$$x (t_1) = 0. \tag{18.86}$$

Conformément à (85), (80) et (86), on a

$$X(t_1)\,x(0)+X(t_1)\int_0^{t_1} Y(\tau)\,u(\tau)\,d\tau = x(t_1)=0. \qquad (18.87)$$

Comme la matrice $X(t)$ est régulière, (87) entraîne la relation

$$x(0) = -\int_0^{t_1} Y(\tau)\,u(\tau)\,d\tau. \qquad (18.88)$$

La relation vectorielle (88) est équivalente aux relations scalaires

$$x_s(0) = -\sum_{l=1}^{r}\int_0^{t_1} Y_{sl}(\tau)\,u_l(\tau)\,d\tau \qquad (s=1,\ \ldots,\ n). \qquad (18.89)$$

En remplaçant $u_l(\tau)$ par la commande optimale (84), ramenons les relations (89) à la forme

$$x_s(0) = -\sum_{l=1}^{r}\int_0^{t_1} Y_{sl}(\tau)\left[\operatorname{sign}\sum_{j=1}^{n} Y_{jl}(\tau)\,\psi_j(0)\right]d\tau \qquad (s=1,\ \ldots,\ n).$$

$$(18.90)$$

Dans les relations (90) les inconnues sont les grandeurs $\psi_1(0), \ldots$ $\ldots, \psi_n(0)$, ainsi que la valeur de l'instant t_1, où le système atteint l'origine des coordonnées.

Les équations (90) sont les équations initiales pour certaines méthodes numériques de recherche de la commande optimale.

§ 19. Problème à extrémités libres. Application du principe du maximum. Conditions de transversalité

Il existe des problèmes dans lesquels l'état final du système est donné sur un ensemble M_1 (plan, ligne, etc.) de l'espace de phase, et non pas sous la forme d'un point fixé de cet espace. Ainsi, par exemple, des problèmes sont possibles dans lesquels à l'instant final t_1 seules les coordonnées de phase du système $x_{\sigma_1}, x_{\sigma_2}, \ldots, x_{\sigma_p}$ présentent de l'intérêt. Les autres $n-p$ coordonnées de phase peuvent avoir à l'instant final t_1 des valeurs arbitraires.

Des problèmes sont également possibles dans lesquels l'état initial $x(t_0)$ n'est pas donné à l'avance et on sait seulement que le point $x(t_0)$ appartient à l'ensemble M_0 de l'espace de phase.

On tombe ainsi sur le problème optimal à extrémités libres dont les solutions obtenues dans la monographie [72] seront exposées dans ce qui suit.

Avant de donner la formulation exacte du problème, il convient de mettre au point les caractéristiques géométriques des ensembles M_0 et M_1 susmentionnés [72]. Soit X un espace de dimension n aux coordonnées orthogonales $x_1, \ldots, x_n$. L'ensemble S des points $x = (x_1, \ldots, x_n)$ vérifiant la relation

$$f(x_1, \ldots, x_n) = 0, \tag{19.1}$$

s'appelle hypersurface de l'espace X, et la relation (1) est dite équation de cette hypersurface. Le point $x \in S$ qui satisfait aux relations

$$\frac{\partial f(x)}{\partial x_1} = \frac{\partial f(x)}{\partial x_2} = \ldots = \frac{\partial f(x)}{\partial x_n} = 0$$

(on suppose que les dérivées $\frac{\partial f}{\partial x_i}$ existent) s'appelle *point singulier* de l'hypersurface S. Ainsi, en un point singulier le vecteur

$$\operatorname{grad} f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$$

est nul. Les points de l'hypersurface S où $\operatorname{grad} f(x) \neq 0$ s'appellent points réguliers.

L'hypersurface définie par l'équation (1) au premier membre continûment dérivable et ne possédant pas de points singuliers s'appelle *hypersurface lisse*. (Toutes les hypersurfaces étudiées dans ce qui suit sont supposées lisses.)

Si l'équation (1) est linéaire, c'est-à-dire si elle est de la forme

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n + b = 0, \tag{19.2}$$

l'absence de points singuliers signifie qu'au moins un des coefficients a_i est différent du zéro. Dans ce cas l'hypersurface, définie par l'équation (2), s'appelle *hyperplan*.

Le vecteur $\operatorname{grad} f(x^0)$ se nomme *vecteur normal* à l'hypersurface S au point x^0. Comme on le voit de (2), dans le cas d'un hyperplan, les vecteurs normaux sont les mêmes en tous les points et ont la forme

$$(a_1, a_2, \ldots, a_n).$$

Tout hyperplan est défini univoquement par un vecteur normal et un point appartenant à ce plan.

Soit S une hypersurface lisse définie par l'équation (1), et x^0 un de ses points. L'hyperplan qui passe par le point x^0 et dont la normale est le vecteur $\operatorname{grad} f(x^0)$ est dit *hyperplan tangent* à l'hypersurface S en x^0. Tout vecteur issu du point x^0 situé dans l'hyperplan tangent s'appelle *vecteur tangent* à l'hypersurface S au point x^0.

Soient maintenant $S_1, \ldots, S_k$ des hypersurfaces lisses données dans l'espace X par les équations

$$\left.\begin{array}{l} f_1(x_1, x_2, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots \\ f_k(x_1, x_2, \ldots, x_n) = 0. \end{array}\right\} \tag{19.3}$$

Si l'on observe la condition suivante: en chaque point $x \in M$ les vecteurs

$$\operatorname{grad} f_1(x), \operatorname{grad} f_2(x), \ldots, \operatorname{grad} f_k(x) \tag{19.4}$$

sont linéairement indépendants, l'intersection M de toutes les hypersurfaces (3) (c'est-à-dire l'ensemble de tous les points $x \in X$ vérifiant à la fois toutes les équations (3)) s'appelle *variété lisse de dimension* $(n - k)$ *dans* X.

Constatons que, par définition, la variété de dimension r dans X est donnée par le système de $n - r$ équations. En particulier, la variété de dimension $(n - 1)$ est donnée par une équation de la forme (3), c'est-à-dire c'est une hypersurface. Les variétés à une dimension sont données par $n - 1$ équations de la forme (3) ; on les appelle également *lignes*.

Si les équations (3) définissant la variété M de dimension $(n - k)$ sont linéaires, la variété M s'appelle *plan de dimension* $(n - k)$ de l'espace X. Les plans à une dimension s'appellent également *lignes droites*.

Soit M une variété lisse de dimension $(n - k)$ définie dans l'espace X par les équations (3), et x un de ses points. Désignons par L_i $(i = 1, \ldots, k)$ l'hyperplan tangent à l'hypersurface $f_i(x_1, x_2, \ldots, x_n) = 0$ au point x. L'intersection des hyperplans $L_1, L_2, \ldots, L_k$ est un plan de dimension $(n - k)$ appelé *plan tangent de la variété* M en x. Le vecteur issu du point x repose dans le plan tangent (c'est-à-dire constitue *un vecteur tangent à la variété* M en x) si et seulement si il est orthogonal à tous les vecteurs (4).

Maintenant on peut donner l'énoncé exact du problème optimal étudié.

Soient M_0 et M_1 des variétés lisses de dimensions arbitraires r_0 et r_1 (mais plus petites que n), situées dans l'espace X (fig. 19.1). On demande de trouver la commande admissible $u(t)$ qui ramène le point représentatif d'une position (non donnée à l'avance) $x^0 \in M_0$ à une position $x^* \in M_1$ et minimise la fonctionnelle donnée.

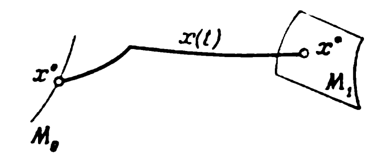


Fig. 19.1

Dans le cas où les deux variétés dégénèrent en points, le problème aux extrémités libres se transforme en un problème aux extrémités fixées.

Remarquons que si nous connaissions les points x^0 et x^*, nous aurions un problème aux extrémités fixées. C'est pourquoi la commande $u(t)$, optimale pour le problème à extrémités libres, est encore optimale pour le cas où les points x^0 et x^* sont considérés comme connus, c'est-à-dire que le principe du maximum (théorème 1, p. 243) est également valable pour le problème à extrémités libres.

Cependant, il faut encore connaître les relations qui permettent de déterminer la position des points x^0 et x^* sur les variétés M_0 et M_1. Ces relations sont les *conditions de transversalité* qui permettent d'écrire $r_0 + r_1$ relations où interviennent les coordonnées des points d'extrémité x^0 et x^*.

Constatons que le nombre de paramètres inconnus (par rapport au problème à extrémités fixées) s'est également accru de $r_0 + r_1$ du fait que la position du point x^0 sur la variété M_0 est définie par r_0 paramètres, alors que celle du point x^* sur la variété M_1 par r_1 paramètres.

Voici les conditions de transversalité. Soient $x^0 \in M_0$, $x^* \in M_1$, des points, et T_0, T_1, les plans tangents des variétés M_0 et M_1 menés par ces points. Les dimensions des plans T_0 et T_1 sont r_0 et r_1 respectivement.

Soient $u(t)$, $x(t)$ $(t_0 \leqslant t \leqslant t_1)$ la solution du problème à extrémités fixées x^0 et x^*, et $\psi(t)$ le vecteur dont l'existence est démontrée par le théorème 1 (p. 243).

La condition de transversalité à l'extrémité droite de la trajectoire $x(t)$, *c'est-à-dire au point* $x(t_1)$, *consiste dans le fait que le vecteur* $\psi(t_1) = (\psi_1(t_1), \psi_2(t_1), \ldots, \psi_n(t_1))$ *est orthogonal au plan* T_1. Autrement dit, tout vecteur $\Theta = (\theta_1, \theta_2, \ldots, \theta_n)$ appartenant au plan T_1 vérifie la relation

$$\langle \psi(t_1), \Theta \rangle = 0. \tag{19.5}$$

La condition de transversalité à l'extrémité gauche de la trajectoire $x(t)$ a un sens analogue (il ne faut que remplacer t_1 et T_1 par t_0 et T_0 respectivement).

Comme dans la relation (5) on ne peut substituer que r_1 vecteurs linéairement indépendants $\Theta_1, \Theta_2, \ldots, \Theta_{r_1}$, situés dans le plan T_1, la condition de transversalité à l'extrémité droite de la trajectoire $x(t)$ donne r_1 relations indépendantes. La condition de transversalité à l'extrémité gauche donne r_0 relations indépendantes.

Ainsi, le théorème suivant a lieu [72].

Théorème. *Soit* $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ *une commande admissible qui ramène le point représentatif d'une position* $x^0 \in M_0$ *en une position* $x^* \in M_1$ *et* $x(t)$, *la trajectoire correspondante issue du point* $x^0 = (x_1, x_2, \ldots, x_n)$. *Pour que* $u(t)$ *et* $x(t)$ *donnent la solution du problème optimal, il faut qu'il existe un vecteur fonction non nul continu* $\psi(t)$ *vérifiant les conditions du théorème* 1 (p. 243) *et de plus, la condition de transversalité à deux extrémités de la trajectoire* $x(t)$.

§ 20. Notion de synthèse régulière dans la théorie des systèmes optimaux

La notion de *synthèse régulière* a été introduite par V. Boltianski [14] pour un système régi par les équations différentielles (17.1)

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r) \qquad (j = 1, \ldots, n),$$

ou par l'équation vectorielle équivalente

$$\frac{dx}{dt} = f(x, u), \tag{20.1}$$

avec $u \in \Omega$ et les dérivées

$$\frac{\partial f_j}{\partial x_i}, \ \frac{\partial f_j}{\partial u_k} \qquad (j, i = 1, \ldots, n; \ k = 1, \ldots, r),$$

supposées continues.

Supposons donnés un ensemble N lisse par morceaux (cf. § 15, p. 231) de dimension $\leqslant n - 1$, des ensembles lisses par morceaux

$$P^0 \subset P^1 \subset P^2 \subset \ldots \subset P^{n-1} \subset P^n = D \tag{20.2}$$

et la fonction $v(x)$ définie dans D et prenant des valeurs dans Ω.

Si les conditions suivantes énoncées par V. Boltianski sont vérifiées, les ensembles (2) et la fonction $v(x)$ réalisent une *synthèse régulière* de l'équation (1) dans le domaine D.

A. L'ensemble P^0 contient le point $a = x^*$ et ne possède pas de points d'accumulation dans l'ensemble ouvert D. Toute composante de l'ensemble $P^i \setminus (P^{i-1} \cup N)$ $(i = 1, 2, \ldots, n)$ est une variété lisse de dimension i dans D ; ces composantes s'appellent cases de dimension i. Les points de l'ensemble P^0 s'appellent cases de dimension nulle. La fonction $v(x)$ est continue et continûment dérivable dans chaque case et peut être prolongée en une fonction continûment dérivable dans le voisinage de la case.

B. Toutes les cases sont rangées en première et deuxième espèces. Les cases de dimension n sont de première espèce, les cases de dimension nulle sont de deuxième espèce.

C. Si σ est une case de dimension i de *première espèce*, par tout point de cette case passe une seule trajectoire de l'équation

$$\frac{dx}{dt} = f(x, v(x)). \tag{20.3}$$

Il existe une case $\Pi(\sigma)$ de dimension $(i - 1)$ telle que toute trajectoire du système (3) qui passe par la case σ, la quitte après un temps fini en butant sous un angle non nul contre la case $\Pi(\sigma)$ et en s'approchant de cette dernière avec une vitesse de phase non nulle. Si σ est une case à *une dimension* de première espèce, elle présente un morceau de trajectoire de phase du système (3) qui s'approche avec une vitesse de phase non nulle d'une certaine case $\Pi(\sigma)$ de dimension nulle.

Si σ est une case de dimension i de *deuxième espèce*, différente du point a, il existe une case $\Sigma(\sigma)$ de dimension $(i + 1)$ de première espèce telle que d'un point quelconque de la case σ est issue une seule trajectoire du système (3) qui passe par la case $\Sigma(\sigma)$; de plus, la fonction $v(x)$ est continue et continûment dérivable sur $\sigma \cup \Sigma(\sigma)$.

D. Les conditions énoncées ci-dessus rendent possible le prolongement de la trajectoire du système (3) d'une case à l'autre: de la case σ à la case $\Pi(\sigma)$, si la case $\Pi(\sigma)$ est de première espèce, et de la case σ à la case $\Sigma(\Pi(\sigma))$, si la case $\Pi(\sigma)$ est de deuxième espèce. Il faut que chacune de ces trajectoires ne passe que par un nombre fini de cases (c'est-à-dire que le « perçage » des cases de deuxième espèce ne s'effectue par chaque trajectoire qu'un nombre fini de fois). De plus, toute trajectoire se termine au point a.

Les trajectoires de ce type sont dites *marquées*.

Ainsi de chaque point de l'ensemble $D \setminus N$ part une seule trajectoire marquée qui mène au point a. Il faut également que de chaque point de l'ensemble N parte une trajectoire (non unique, si possible) du système (3), qui mène au point a et qu'on appelle également marquée.

E. Toutes les trajectoires marquées vérifient le principe du maximum.

F. La valeur de la fonctionnelle Q (cf. (17.3)), calculée le long des trajectoires marquées (qui se terminent au point a), est une fonction continue du point initial $x(t_0)$.

En particulier, si plusieurs trajectoires marquées sont issues du point $x(t_0) \in N$ la valeur de la fonctionnelle Q est la même pour toutes ces trajectoires.

Exemple de synthèse régulière. Problème de la réponse en temps minimal dans un système linéaire. Tous les exemples connus de synthèse d'une commande à temps optimal dans les systèmes linéaires sont des cas particuliers de synthèse régulière.

Montrons-le sur l'exemple relativement simple du système linéaire du § 15, p. 227.

Les équations du mouvement de ce système s'écrivent

$$\frac{dx_1}{dt} = x_2, \quad \frac{dx_2}{dt} = u. \tag{20.4}$$

La commande u subit la contrainte

$$-1 \leqslant u \leqslant 1. \tag{20.5}$$

On demande de ramener le système de l'état initial donné $x(t_0)$ à l'état $x = 0$ en un temps minimal.

L'intervalle de temps pendant lequel on réussit à recaler le système à l'état $x = 0$ dépend essentiellement des contraintes imposées à la commande u: $|u| \leqslant 1$.

Conformément à (1) et (4), dans notre exemple,

$$f_1(x, u) = x_2, \qquad f_2(x, u) = u. \tag{20.6}$$

D'après (17.21), la fonction H se met ici sous la forme

$$H = \psi_1 x_2 + \psi_2 u, \tag{20.7}$$

où, en vertu de (17.23), ψ_1 et ψ_2 vérifient les équations différentielles

$$\frac{d\psi_1}{dt} = 0, \quad \frac{d\psi_2}{dt} = -\psi_1. \tag{20.8}$$

Il ressort de l'expression (7) que pour un instant quelconque t, sous la contrainte (5), la valeur maximale de la fonction $H(\psi(t), x(t), u)$ de la variable u, où $|u| \leqslant 1$, est donnée par la commande de

$$u(t) = \operatorname{sign} \psi_2(t). \tag{20.9}$$

Ainsi, d'après (9), la commande optimale est

$$u(t) = 1 \text{ pour } \psi_2(t) > 0, \qquad u(t) = -1 \text{ pour } \psi_2(t) < 0.$$

D'après (8), les fonctions $\psi_1(t)$ et $\psi_2(t)$ sont de la forme

$$\psi_1(t) = \psi_1(t_0), \ \psi_2(t) = -\psi_1(t_0)(t - t_0) + \psi_2(t_0), \tag{20.10}$$

où par $\psi_1(t_0)$ et $\psi_2(t_0)$ on désigne les valeurs initiales des fonctions $\psi_1(t)$ et $\psi_2(t)$ respectivement.

Si les valeurs initiales $\psi_1(t_0)$ et $\psi_2(t_0)$ étaient connues, alors, en vertu de (9) et (10) le problème de synthèse de la commande optimale serait résolu. La recherche de la valeur initiale $\psi(t_0)$ du vecteur fonction $\psi(t)$ constitue la plus grande difficulté du problème général de synthèse de la commande optimale par rapport au principe du maximum.

Dans notre exemple bien simple on réussit à résoudre le problème grâce à la possibilité de définir explicitement dans le plan de phase x_1x_2 la courbe $x_2 = \varphi(x_1)$ sur laquelle la commande u (dont le module d'après (9) reste constant: $|u| = 1$) doit changer de signe. La courbe $x_2 = \varphi(x_1)$ s'appelle *ligne de commutation de la commande.*

Puisque d'après (9)

$$u(t) = \operatorname{sign}\psi_2(t) = \operatorname{sign}[-\psi_1(t_0)(t-t_0) + \psi_2(t_0)], \tag{20.11}$$

et la fonction $\psi_2 = -\psi_1(t_0)(t-t_0) + \psi_2(t_0)$ ne change son signe qu'une seule fois à l'instant $t_c = \frac{\psi_2(t_0)}{\psi_1(t_0)} + t_0$ la commande optimale $u(t)$ change son signe pas plus d'une fois dans l'intervalle de temps $t_0 \leqslant t \leqslant t_1$, avec t_1 l'instant où le point représentatif rejoint l'origine des coordonnées.

Ainsi la commande optimale $u = u(t)$ est une fonction continue par morceaux (son module est $|u(t)| = 1$), qui compte au plus deux intervalles de permanence.

Cette propriété de la commande optimale $u(t)$ permet dans notre exemple de construire une famille de trajectoires optimales dans le plan de phase x_1x_2.

Dans l'intervalle de temps où la commande $u = 1$, les équations (4) se mettent sous la forme

$$\frac{dx_1}{dt} = x_2, \quad \frac{dx_2}{dt} = 1. \tag{20.12}$$

d'où l'on déduit

$$\frac{dx_1}{dx_2} = x_2.$$

En intégrant cette équation, on obtient la famille des paraboles

$$x_1 = \frac{x_2^2}{2} + C_1, \tag{20.13}$$

dépendant du paramètre C_1. Avec $C_1 = 0$ on obtient la parabole

$$x_1 = \frac{x_2^2}{2}, \tag{20.14}$$

qui passe par l'origine des coordonnées. L'arc AO de la parabole (14) est représenté sur la figure 20.1. Les lignes K_iB_i ($i = 1, 2, \ldots$) sont des arcs des paraboles (13) qui correspondent aux différentes valeurs du paramètre C_1.

D'une façon analogue, avec $u = -1$, les équations (4) deviennent

$$\frac{dx_1}{dt} = x_2, \quad \frac{dx_2}{dt} = -1, \tag{20.15}$$

d'où l'on tire que

$$\frac{dx_1}{dx_2} = -x_2.$$

En intégrant cette équation, on trouve la famille des paraboles

$$x_1 = -\frac{x_2^2}{2} + C_2, \tag{20.16}$$

qui dépend du paramètre C_2. Avec $C_2 = 0$ on obtient la parabole

$$x_1 = -\frac{x_2^2}{2}, \tag{20.17}$$

qui passe par l'origine des coordonnées. L'arc BO de la parabole (17) est représenté sur la figure 20.1. Les lignes L_iA_i sont des arcs des paraboles (16) qui correspondent à de différentes valeurs du paramètre C_2.

Désignons maintenant par Γ la ligne composée d'arcs AO et OB. La ligne Γ divise le plan de phase x_1x_2 en deux domaines D_1 et D_2 situés respectivement au-dessus et au-dessous de la ligne Γ.

Si à l'instant initial t_0 le point représentatif se trouve dans le domaine D_1, par exemple au point L_i, il faut poser $u = -1$. Alors, le point représentatif

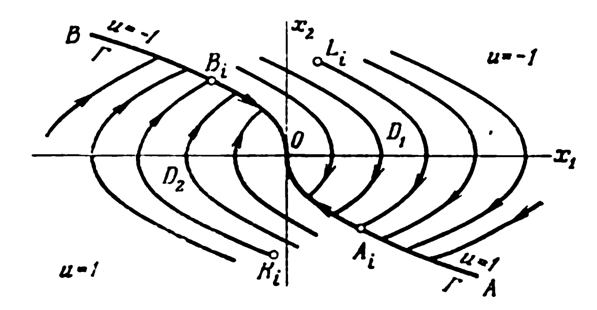


Fig. 20.1

suit l'arc L_iA_i de la parabole (16) qui passe par le point $x(t_0) = L_i$. A l'instant où le point représentatif rejoint le point A_i, il faut commuter la commande pour obtenir $u = 1$. Le point représentatif poursuit alors son mouvement suivant l'arc A_iO, qui le mène à l'origine des coordonnées. La courbe L_iA_iO, composée d'arcs L_iA_i et A_iO, est la *trajectoire optimale* qui correspond à l'état initial $x(t_0) = L_i$. Suivant le théorème de l'unicité de la trajectoire optimale du problème de la réponse en temps minimal dans les systèmes linéaires [72], la durée du mouvement suivant la trajectoire L_iA_iO est minimale par rapport à la durée du mouvement suivant une autre trajectoire quelconque qui mène du point L_i à l'origine des coordonnées.

D'une façon analogue, si à l'instant initial le point représentatif se trouve dans le domaine D_2, par exemple au point K_i, il faut poser $u = 1$. Le point représentatif suit alors l'arc K_iB_i de la parabole (13). Au point B_i il faut commuter la commande pour obtenir $u = -1$. Le mouvement ultérieur se poursuit suivant l'arc B_iO qui mène le point représentatif à l'origine des coordonnées. Ainsi, pour l'état initial $x(t_0) = K_i$, la trajectoire optimale est K_iB_iO.

Comme nous l'avons dit plus haut, nous avons démontré pour le problème de la réponse en temps minimal dans les systèmes linéaires le théorème de l'unicité de la trajectoire optimale [72] mise en évidence par le principe du maximum.

Dans notre exemple, la démonstration de l'*unicité* de la trajectoire optimale obtenue ci-dessus peut être immédiate [15]. Désignons par t_1 l'instant où le point représentatif atteint l'origine des coordonnées. La durée du mouvement suivant la trajectoire L_iA_iO est alors $t_1 - t_0$. Désignons par t_c l'instant de commutation de la commande.

Supposons qu'il existe une autre commande $\tilde{u}(t)$, où $|\tilde{u}(t)| \leqslant 1$, qui amène le système à l'origine des coordonnées en un temps $t_1^* - t_0$, où $t_1^* < t_1$.

Désignons toujours la trajectoire associée à la commande (9) par $x(t)$ et à la commande $\tilde{u}(t)$ par $\tilde{x}(t)$. On a ainsi les relations suivantes

$$x_1(t_1) = 0, \qquad x_2(t_1) = 0,$$
$$\tilde{x}_1(t_1^*) = 0, \qquad \tilde{x}_2(t_1^*) = 0.$$

Les trajectoires $x(t)$ et $\tilde{x}(t)$ sont définies par les équations du mouvement

$$\dot{x}_1(t) = x_2(t), \qquad \dot{x}_2(t) = u(t) \qquad (t_0 \leqslant t \leqslant t_1)$$

et, respectivement,

$$\dot{\tilde{x}}_1(t) = \tilde{x}_2(t), \; \dot{\tilde{x}}_2(t) = \tilde{u}(t) \qquad (t_0 \leqslant t \leqslant t_1^*).$$

Introduisons maintenant les notations

$$\Phi(t) = -x_1(t) + x_2(t)(t - t_c),$$
$$\Psi(t) = -\tilde{x}_1(t) + \tilde{x}_2(t)(t - t_c).$$

Puisque

$$x_1(t_0) = \tilde{x}_1(t_0), \qquad x_2(t_0) = \tilde{x}_2(t_0),$$

il vient

$$\Phi(t_0) = \Psi(t_0).$$

Les relations ci-dessus entraînent

$$\Phi(t_1) = 0, \qquad \Psi(t_1^*) = 0.$$

Les dérivées par rapport à t de $\Phi(t)$ et $\Psi(t)$ sont

$$\frac{d\Phi}{dt} = \frac{\partial\Phi}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial\Phi}{\partial x_2}\frac{dx_2}{dt} + \frac{\partial\Phi}{\partial t} = (t - t_c)\,u,$$

$$\frac{d\Psi}{dt} = \frac{\partial\Psi}{\partial \tilde{x}_1}\frac{d\tilde{x}_1}{dt} + \frac{\partial\Psi}{\partial \tilde{x}_2}\frac{d\tilde{x}_2}{dt} + \frac{\partial\Psi}{\partial t} = (t - t_c)\,\tilde{u}.$$

Comme sur la trajectoire $L_i A_i O$ la commande $u(t)$ vérifie les conditions

$$u(t) = -1 \text{ pour } t \in [t_0, t_c), \; u(t) = 1 \text{ pour } t \in [t_c, t_1],$$

alors,

$$(t - t_c)\,u(t) = |t - t_c| \qquad (t_0 \leqslant t \leqslant t_1).$$

On voit aisément que toute autre commande $\tilde{u}(t)$, où $|\tilde{u}(t)| \leqslant 1$, donne lieu à l'inégalité

$$(t - t_c)\,\tilde{u}(t) \leqslant |t - t_c| \qquad (t_0 \leqslant t \leqslant t_1^*).$$

Des relations obtenues il suit

$$\dot{\Phi}(t) \geqslant \dot{\Psi}(t), \qquad t \in [t_0, t_1^*],$$

d'où l'on déduit

$$\int_{t_0}^{t_1^*} \dot{\Phi}\,dt \geqslant \int_{t_0}^{t_1^*} \dot{\Psi}\,dt.$$

Ainsi

$$\Phi(t_1^*) - \Phi(t_0) \geqslant \Psi(t_1^*) - \Psi(t_0),$$

et, puisque $\Phi(t_0) = \Psi(t_0)$, et, de plus, $\Psi(t_1^*) = 0$, on a

$$\Phi(t_1^*) \geqslant 0.$$

Par ailleurs, puisque nous avons supposé que $t_1 > t_1^*$, compte tenu du fait que $\Phi(t_1) = 0$, on peut obtenir la relation

$$-\Phi(t_1^*) = \Phi(t_1) - \Phi(t_1^*) = \int_{t_1^*}^{t_1} \dot{\Phi}\, dt = \int_{t_1^*}^{t_1} |t - t_c|\, dt > 0,$$

d'où il résulte que

$$\Phi(t_1^*) < 0.$$

On aboutit ainsi à une contradiction. Par conséquent, il est impossible que t_1^* soit inférieur à t_1.

Supposons maintenant que $t_1^* = t_1$ mais qu'il existe une commande $\tilde{u}(t)$ différente de $u(t)$ qui amène le système à l'instant t_1 à l'origine des coordonnées.

Puisque $\tilde{u}(t) \not\equiv u(t)$, $t \in [t_0, t_1]$, ce qui vient d'être dit entraîne que

$$\int_{t_0}^{t_1} \dot{\Phi}\, dt > \int_{t_0}^{t_1} \dot{\Psi}\, dt,$$

soit

$$\Phi(t_1) - \Phi(t_0) > \Psi(t_1) - \Psi(t_0).$$

Comme $\Phi(t_0) = \Psi(t_0)$, $\Phi(t_1) = 0$, on a

$$0 > -\tilde{x}_1(t_1) + \tilde{x}_2(t_1)(t_1 - t_c),$$

ce qui contredit l'hypothèse suivant laquelle

$$\begin{bmatrix} \tilde{x}_1(t_1) \\ \tilde{x}_2(t_1) \end{bmatrix} = 0.$$

Par suite, il n'existe pas de commande différente de $u(t)$ susceptible de recaler le système à l'origine à l'instant t_1.

Ainsi nous avons démontré l'unicité de la commande optimale (9) (sous la contrainte $|u| \leqslant 1$) et l'unicité de la trajectoire optimale associée à cette commande pour l'état initial donné.

Cherchons maintenant la fonction de Bellman pour le problème considéré

$$\omega(x(t_0)) = -T(x(t_0)). \tag{20.18}$$

Ici

$$T(x(t_0)) = t_1 - t_0 \tag{20.19}$$

est la durée du mouvement du point représentatif suivant la trajectoire optimale de la position initiale $x(t_0)$ à l'origine des coordonnées $x = 0$.

Supposons qu'à l'instant initial $t = t_0$ le point représentatif se trouve au point L_i du plan x_1x_2, c'est-à-dire que $x_1(t_0)$ et $x_2(t_0)$ sont les coordonnées du point L_i. En vertu de (16), l'équation de la parabole dont l'arc est la ligne L_iA_i a pour expression

$$x_1 = -\frac{x_2^2}{2} + x_1(t_0) + \frac{x_2^2(t_0)}{2}. \tag{20.20}$$

Le point A_i de commutation de la commande est le point d'intersection des paraboles (20) et (14). L'ordonnée du point A_i vérifie la relation

$$(x_{A_i})_2^2 = x_1(t_0) + \frac{x_2^2(t_0)}{2}.$$

L'arc AO reposant au-dessous de l'axe des abscisses, il faut poser

$$(x_{A_i})_2 = -\left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2}. \tag{20.21}$$

Le mouvement du point représentatif suivant l'arc L_iA_i donne lieu aux équations (15). La deuxième de ces équations est de la forme

$$\frac{dx_2}{dt} = -1,$$

d'où, compte tenu du fait que les coordonnées du point L_i sont désignées ici par $x_1(t_0)$ et $x_2(t_0)$, on a

$$\int_{x_2(t_0)}^{(x_{A_i})_2} dx_2 = -\int_{t_0}^{t_c} dt, \tag{20.22}$$

où t_c est l'instant de commutation de la commande.

Il résulte de (22) que

$$(x_{A_i})_2 - x_2(t_0) = t_0 - t_c. \tag{20.23}$$

D'après (12), le mouvement du point représentatif suivant l'arc A_iO donne lieu à l'équation

$$\frac{dx_2}{dt} = 1,$$

d'où

$$\int_{(x_{A_i})_2}^{0} dx_2 = \int_{t_c}^{t_1} dt, \tag{20.24}$$

avec t_1, l'instant où le point représentatif rejoint l'origine des coordonnées. L'expression (24) entraîne

$$-(x_{A_i})_2 = t_1 - t_c. \tag{20.25}$$

Il résulte de (23) et (25) que

$$T(x(t_0)) = t_1 - t_0 = x_2(t_0) - 2(x_{A_i})_2. \tag{20.26}$$

Conformément à (21), l'expression (26) se met sous la forme

$$T(x(t_0)) = x_2(t_0) + 2\left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2} \quad \text{pour } x(t_0) \in D_1. \tag{20.27}$$

L'expression (27) a lieu lorsque le point $x(t_0)$ se situe dans le domaine D_1, c'est-à-dire en n'importe quel point du plan de phase au-dessus de la ligne AOB.

Toutefois, on voit aisément que l'expression (27) reste valable, si le point $x(t_0)$ repose sur la ligne AO. En effet, dans ce cas, $t_c = t_0$, $x_2(t_0) = (x_{A_i})_2$, et pour les points de la ligne AO l'expression (26) devient

$$T(x(t_0)) = -(x_{A_i})_2 = |(x_{A_i})_2| \text{ pour } x(t_0) \in AO,$$

ce qui coïncide avec l'expression (25) pour la durée du mouvement du point représentatif suivant l'arc A_iO.

D'une façon analogue on peut trouver $T(x(t_0))$ pour le cas où le point $x(t_0)$ repose dans le domaine D_2 et sur la ligne BO.

Dans le cas où le point $x(t_0)$ se situe dans le domaine D_2, c'est-à-dire en un point quelconque du plan de phase au-dessous de la ligne AOB, on a

$$T(x(t_0)) = -x_2(t_0) + 2\left[-x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2} \quad \text{pour } x(t_0) \in D_2. \qquad (20.28)$$

Si à l'instant initial le point représentatif se trouve au point B_i de la ligne BO, le temps nécessaire pour recaler le système à l'origine des coordonnées est

$$T(x(t_0)) = (x_{B_i})_2 \quad \text{pour } x(t_0) \in BO. \qquad (20.29)$$

Les formules (28) et (27) amènent que pour les points initiaux symétriques K_i et L_i le temps d'amenée à l'origine des coordonnées est naturellement le même.

On voit aisément qu'en points de la ligne AO, les fonctions (27) et (28) prennent les mêmes valeurs. En effet, puisque AO est un arc de la parabole (14)

$$x_1 = \frac{x_2^2}{2},$$

aux points de la ligne AO la fonction (27) prend une valeur qui peut s'écrire, compte tenu du fait que la ligne AO repose au-dessous de l'axe des abscisses,

$$[T(x(t_0))]_1 = -|x_2(t_0)| + 2|x_2(t_0)| = |x_2(t_0)| \quad (x(t_0) \in AO). \qquad (20.30)$$

Aux points de la ligne AO la fonction (28) prend la valeur

$$[T(x(t_0))]_2 = -x_2(t_0) = |x_2(t_0)| \quad (x(t_0) \in AO). \qquad (20.31)$$

Les valeurs (30) et (31) coïncident. D'une façon analogue on peut montrer que les valeurs des fonctions (27) et (28) coïncident également aux points de la ligne BO; la valeur qu'ils prennent est égale à $x_2(t_0)$.

Ainsi, la fonction $T(x(t_0))$ déterminée dans les domaines D_1 et D_2 respectivement par les expressions (27) et (28) conserve sa *continuité* dans tout le plan de phase x_1x_2.

Les dérivées partielles de la fonction (27) par rapport à ses variables sont

$$\left.\begin{aligned} \frac{\partial T}{\partial x_1(t_0)} &= \left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{-1/2}, \\ \frac{\partial T}{\partial x_2(t_0)} &= 1 + x_2(t_0)\left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{-1/2} \quad (x(t_0) \in D_1). \end{aligned}\right\} \qquad (20.32)$$

D'une façon analogue les dérivées partielles de la fonction (28) sont

$$\left.\begin{aligned} \frac{\partial T}{\partial x_1(t_0)} &= -\left[-x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{-1/2}, \\ \frac{\partial T}{\partial x_2(t_0)} &= -1 + x_2(t_0)\left[-x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{-1/2} (x(t_0) \in D_2). \end{aligned}\right\} \qquad (20.33)$$

Cherchons les valeurs que prennent les fonctions (32) et (33) au point A_i, c'est-à-dire posons que

$$x_1(t_0) = (x_{A_i})_1, \qquad x_2(t_0) = (x_{A_i})_2.$$

Comme d'après (21)

$$(x_{A_l})_2 = -\left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2},$$

les valeurs des fonctions (32) en A_l sont

$$\left.\begin{aligned} \left.\frac{\partial T}{\partial x_1(t_0)}\right|_{A_l \in \overline{D}_1} &= -\frac{1}{(x_{A_l})_2}, \\ \left.\frac{\partial T}{\partial x_2(t_0)}\right|_{A_l \in \overline{D}_1} &= 1 + \frac{(x_{A_l})_2}{-(x_{A_l})_2} = 0. \end{aligned}\right\} \qquad (20.34)$$

Les valeurs des fonctions (33) en A_l où, comme sur toute la ligne AO, $x_1 = \frac{x_2^2}{2}$, sont

$$\left.\frac{\partial T}{\partial x_1(t_0)}\right|_{A_l \in \overline{D}_2} = -\infty, \qquad \left.\frac{\partial T}{\partial x_2(t_0)}\right|_{A_l \in \overline{D}_2} = -\infty. \qquad (20.35)$$

(34) et (35) montrent que les valeurs des fonctions (32) et (33) sur la ligne $AO\left(\text{arc de la courbe } x_1 = \frac{x_2^2}{2}\right)$ sont différentes. Sur la ligne BO $\left(\text{arc de la courbe } x_1 = -\frac{x_2^2}{2}\right)$ les valeurs des fonctions (32) et (33) sont également différentes.

Ainsi, aux points de la ligne AOB la fonction $T(x(t_0))$ et, par conséquent, la fonction $\omega(x(t_0)) = -T(x(t_0))$ ne possèdent pas de dérivées partielles par rapport à leurs variables $x_1(t_0)$ et $x_2(t_0)$.

On voit sans peine que les trajectoires *optimales* construites dans notre exemple (fig. 20.1) vérifient les conditions de la *synthèse régulière* énoncées dans ce qui précède (p. 269).

Constatons de plus que dans l'exemple concerné, l'ensemble $P^0 = 0$, c'est-à-dire l'ensemble P^0 ne contient que le point $x = 0$.

L'ensemble $P^1 = \Gamma$, où par Γ on désigne la ligne composée de AO et OB (c'est-à-dire la ligne de commutation de la commande).

L'ensemble $P^2 = D$, où D est le plan de phase x_1x_2 tout entier.

L'ensemble N de notre exemple est vide. Ainsi, ici $P^{n-1} \cup N = P^1$, c'est-à-dire la réunion des ensembles P^{n-1} et N est la ligne AOB (ligne de commutation de la commande Γ).

Les lignes AO et OB constituent *deux cases à une dimension de première espèce*. Les domaines D_1 et D_2 qui s'obtiennent en divisant le plan de phase par la ligne Γ forment *deux cases de dimension deux de première espèce*.

Notons encore quelques propriétés des trajectoires optimales, suffisamment évidentes dans l'exemple considéré.

Sur la figure 20.2 on désigne par $x(t_0)$ le point où se trouvait le système à l'instant t_0; la trajectoire optimale représentée mène du point $x(t_0)$ à l'origine des coordonnées. La position du point A_l (où la commande est commutée) sur la courbe AO dépend de l'état

initial $x(t_0)$ du système. Comme les points de la courbe AO vérifient la relation (14)

$$(x_{A_i})_1 = \frac{(x_{A_i})_2^2}{2},$$

en désignant

$$\xi_1 = (x_{A_i})_2, \tag{20.36}$$

on peut mettre les coordonnées du point A_i dans le plan x_1x_2 sous la forme

$$\frac{\xi_1^2}{2}, \ \xi_1.$$

Ainsi, la position du point A_i (qui appartient à la ligne AO, c'est-à-dire à la case de dimension un de première espèce) n'est dé-

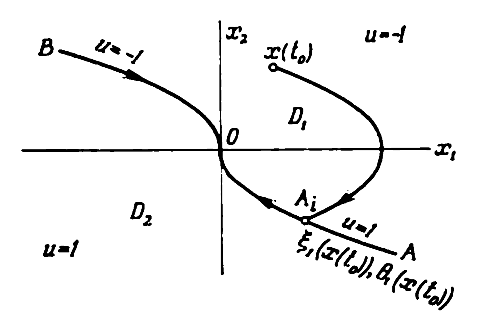


Fig. 20.2

terminée que par un seul paramètre ξ_1 qui est fonction de l'état initial du système

$$\xi_1 = \xi_1(x(t_0)). \tag{20.37}$$

D'après (21) et (36),

$$\xi_1(x(t_0)) = -\left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2}. \tag{20.38}$$

De même, l'intervalle de temps

$$\theta_1 = t_c - t_0, \tag{20.39}$$

pendant lequel le système passe du point $x(t_0)$ en A_i est une fonction de l'état initial du système $x(t_0)$

$$\theta_1 = \theta_1(x(t_0)). \tag{20.40}$$

D'après (23),

$$\theta_1(x(t_0)) = -(x_{A_i})_2 + x_2(t_0).$$

En vertu de (21), cette expression devient

$$\theta_1(x(t_0)) = x_2(t_0) + \left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2}. \tag{20.41}$$

Comme il ressort de (38) et (41), les fonctions $\xi_1(x(t_0))$ et $\theta_1(x(t_0))$ sont des fonctions continues et continûment dérivables de leurs variables $x_1(t_0)$ et $x_2(t_0)$ (c'est-à-dire des coordonnées du vecteur $x(t_0)$).

Puisque le point initial $x(t_0)$ peut être choisi parmi les points quelconques de l'arc x de la trajectoire optimale dans le domaine D_1, les expressions (38) et (41) peuvent s'écrire

$$\xi_1(x) = -\left(x_1 + \frac{1}{2}x_2^2\right)^{1/2}, \qquad (20.42)$$

$$\theta_1(x) = x_2 + \left(x_1 + \frac{1}{2}x_2^2\right)^{1/2}. \qquad (20.43)$$

Considérons maintenant le *mouvement à rebours* du point représentatif qui se déplace du point A_l vers le point $x(t_0)$. Ce mouvement s'obtient en inversant le temps t. Suivant (15), le mouvement à rebours est décrit par les équations différentielles

$$\frac{dy_1}{dt} = -y_2, \quad \frac{dy_2}{dt} = 1. \qquad (20.44)$$

Voici la solution des équations (44):

$$y_1(t) = y_1(0) - y_2(0)\,t - \frac{t^2}{2}, \quad y_2(t) = y_2(0) + t. \qquad (20.45)$$

Comme dans le mouvement à rebours le point initial $y(0)$ se situe sur la ligne AO, on a, d'après (14),

$$y_1(0) = \frac{y_2^2(0)}{2}. \qquad (20.46)$$

C'est pourquoi la solution (45) peut être ramenée à la forme

$$y_1(t, \xi) = \frac{\xi^2}{2} - \xi t - \frac{t^2}{2}, \quad y_2(t, \xi) = \xi + t, \qquad (20.47)$$

où par ξ on désigne

$$\xi = y_2(0). \qquad (20.48)$$

Etant donné que

$$\begin{vmatrix} \frac{\partial y_1}{\partial \xi} & \frac{\partial y_1}{\partial t} \\ \frac{\partial y_2}{\partial \xi} & \frac{\partial y_2}{\partial t} \end{vmatrix} = \begin{vmatrix} \xi - t & -(\xi + t) \\ 1 & 1 \end{vmatrix} = 2\xi \neq 0, \qquad (20.49)$$

l'équation (47) peut être résolue par rapport à ξ et à t.

Remarquons encore que si $x(t_0)$ est l'état initial dans le mouvement de départ (depuis le point $x(t_0)$ vers le point A_l), en portant

dans (47) au lieu de ξ et t les variables

$$\xi = \xi_1(x(t_0)) = -\left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2},$$

$$t = \theta_1(x(t_0)) = x_2(t_0) + \left[x_1(t_0) + \frac{1}{2}x_2^2(t_0)\right]^{1/2},$$

on obtient

$$y_1(\theta_1(x(t_0)), \quad \xi_1(x(t_0))) = x_1(t_0),$$

$$y_2(\theta_1(x(t_0)), \quad \xi_1(x(t_0))) = x_2(t_0),$$

c'est-à-dire si le mouvement à rebours commence (comme il s'ensuit de (36)) au point A_t, après un intervalle de temps θ_1, le point représentatif rejoint naturellement $x(t_0)$.

§ 21. Condition suffisante d'optimalité sous la forme du principe du maximum

Comme nous l'avons dit plus haut, le principe du maximum est une condition *nécessaire* d'optimalité. Par conséquent, le principe du maximum permet de mettre en évidence les trajectoires qui peuvent être optimales.

Pour les *systèmes linéaires du problème de la réponse en temps minimal* on a démontré [72] le théorème d'existence d'une trajectoire optimale et le théorème d'unicité d'une trajectoire mise en évidence par le principe du maximum. Ainsi, dans ce cas, le principe du maximum définit d'une façon *univoque* la trajectoire qui peut être optimale; cette trajectoire est précisément la seule trajectoire optimale qui relie deux points donnés dans l'espace de phase.

Pour les systèmes non linéaires (ainsi que pour les systèmes linéaires des problèmes qui diffèrent du problème de la réponse en temps minimal), la question si la synthèse de la commande, synthèse réalisée sur la base du principe du maximum, aboutit aux trajectoires optimales se résout à l'aide du théorème démontré par V. Boltianski dans [14] sur les *conditions suffisantes d'optimalité*. En général ce théorème permet d'affirmer que la synthèse réalisée sur la base du principe du maximum aboutit réellement aux trajectoires optimales.

Dans le théorème concerné, la condition principale est de réaliser la synthèse régulière de l'équation différentielle vectorielle (20.1) régissant le système de commande. Notons à ce propos que les conditions *A*-*E* (p. 269) de la synthèse régulière n'imposent en fait aucune contrainte au système décrit par cette dernière équation et postulent des circonstances usuelles qui se manifestent lors de la réalisation de la synthèse. Le théorème prouve que *si les conditions A-E sont remplies*, la synthèse réalisée sur la base du principe du

maximum conduit réellement aux trajectoires optimales. En ce sens, le principe du maximum est une *condition suffisante d'optimalité.*

Théorème sur les conditions suffisantes d'optimalité sous la forme du principe du maximum [14]. *Si dans un ensemble D on réalise la synthèse régulière de l'équation* (20.1) (*sous l'hypothèse qu'il existe des dérivées continues* $\frac{\partial f_i}{\partial x_j}$, $\frac{\partial f_i}{\partial u_k}$ *et de la validité de la relation* $f_0(x, u) > 0$) *toutes les trajectoires marquées sont optimales* (*dans le domaine D*). *En ce sens, le principe du maximum est une condition suffisante d'optimalité.*

Démonstration. La démonstration que nous allons exposer a été donnée dans l'article de V. Boltianski [14] publié en 1964. Nous nous bornerons ici au cas où dans l'expression (17.3) la fonction

$$f_0(x, u) \equiv 1,$$

ce qui a lieu dans le problème de la réponse en temps minimal.

Dans ce cas la fonctionnelle Q est la durée du mouvement du point représentatif du point $x(t_0)$ au point a.

Désignons par $-\omega(x)$ la durée du mouvement du point représentatif le long de la trajectoire marquée du point x au point a.

Introduisons la notation M pour l'ensemble $P^{n-1} \cup N$.

Si l'on prouve que $\omega(x)$ est la fonction de Bellman où M est un ensemble singulier, le théorème à démontrer se déduira du théorème démontré au § 15, p. 232.

Donc, il suffit de prouver que la fonction $\omega(x)$, introduite ici, est dérivable sur l'ensemble $D \setminus M$ et vérifie l'équation de Bellman (15.7)

$$\max_{u \in \Omega} \left[\sum_{i=1}^{n} \frac{\partial \omega(x)}{\partial x_i} f_i(x, u) \right] = 1.$$

Soit x un point arbitraire qui appartient à une certaine case σ de dimension n et supposons que la trajectoire du système (20.3)

$$\frac{dx}{dt} = f(x, v(x)),$$

issue à l'instant t_0 du point x indiqué, passe à l'instant $t_0 + \theta_1(x)$ par le point $\xi_1(x)$ de la case $\Pi(\sigma)$.

Il suit des théorèmes généraux sur la dérivabilité des solutions des équations différentielles par rapport aux paramètres que les fonctions $\xi_1(x)$ et $\theta_1(x)$ sont des fonctions continûment dérivables de x.

Pour l'exemple de système linéaire étudié au § 20, la forme explicite des fonctions $\xi_1(x)$ et $\theta_1(x)$ est déterminée par les expressions (20.42) et (20.43). Dans le cas général, ces fonctions se calculent d'après la solution de l'équation différentielle qui décrit le mouve-

ment à rebours

$$\dot{y} = -f(y, v(y)). \tag{21.1}$$

La solution de l'équation (1) est

$$y = y(t, \xi), \tag{21.2}$$

où $\xi = \xi_1(x)$, $t = \theta_1(x)$. L'équation

$$y(t, \xi) = x \tag{21.3}$$

est résoluble d'une façon univoque pour des x voisins de $x(t_0)$ car le déterminant fonctionnel correspondant est différent de zéro. (Dans l'exemple considéré au § 20, les équations du mouvement à rebours sont de la forme (20.44) et le déterminant fonctionnel mentionné, de la forme (20.49).) Les fonctions $\xi_1(x)$ et $\theta_1(x)$ trouvées à partir de (3) sont continûment dérivables par rapport à x.

Après le point $\xi_1(x)$ la trajectoire passe par la case $\Pi(\sigma)$ ou $\Sigma(\Pi(\sigma))$. D'une façon analogue on établit que le point $\xi_2(x)$ où la trajectoire quitte la case $\Pi(\sigma)$ ou $\Sigma(\Pi(\sigma))$, et le temps $\theta_2(x)$ du mouvement dans cette case sont des fonctions dérivables par rapport à $\xi_1(x)$, et donc par rapport à x.

Le temps total $-\omega(x) = \theta_1(x) + \theta_2(x) + \ldots$ du mouvement suivant la trajectoire *marquée* du point x au point a est ainsi (à l'intérieur de la case σ) une fonction continûment dérivable du point x.

Ce qui vient d'être dit entraîne que la fonction $\omega(x)$ *est continûment dérivable sur l'ensemble* $D \setminus M$.

Maintenant il faut montrer que la fonction $\omega(x)$ vérifie sur l'ensemble $D \setminus M$ l'équation de Bellman (15.7).

Soit $x(t_0) \in D \setminus M$. Désignons par $x(t)$ la trajectoire marquée issue à l'instant t_0 du point $x(t_0)$ et qui à l'instant t_1 rejoint le point a.

Désignons par S l'ensemble de tous les points x qui satisfont à la condition

$$\omega(x) = \omega(x(t_0)).$$

Au voisinage du point $x(t_0)$ l'ensemble S forme une hypersurface lisse dans D avec le vecteur

$$\operatorname{grad} \omega(x(t_0)) = \left(\frac{\partial \omega(x(t_0))}{\partial x_1}, \ldots, \frac{\partial \omega(x(t_0))}{\partial x_n} \right)$$

qui lui est normal.

La fonction $\omega(x(t_0)) = -(t_1 - t_0)$, où $t_1 - t_0$ est la durée du mouvement suivant la trajectoire marquée du point $x(t_0)$ au point a. Pour tout point courant $x(t)$ de la trajectoire *marquée*, la fonction $\omega(x(t)) = -(t_1 - t)$, ce qui conduit à $\frac{d\omega(x(t))}{dt} = 1$. Puisqu'en

vertu de (20.3)

$$\left[\frac{d\omega(x(t))}{dt}\right]_{x(t)=x(t_0)} = \sum_{i=1}^{n} \frac{\partial\omega(x(t_0))}{\partial x_i} f_i(x(t_0),\ v(x(t_0))),$$

il vient

$$\sum_{i=1}^{n} \frac{\partial\omega(x(t_0))}{\partial x_i} f_i(x(t_0),\ v(x(t_0))) = 1 \tag{21.4}$$

ou

$$\langle \text{grad}\ \omega(x(t_0)),\ f(x(t_0),\ v(x(t_0)))\rangle = 1. \tag{21.5}$$

Il résulte de (5) que

$$\text{grad}\ \omega(x(t_0)) \neq 0. \tag{21.6}$$

D'après la condition E de la synthèse régulière, la trajectoire marquée vérifie le principe du maximum. Désignons par

$$\psi(t) = (\psi_1(t),\ \ldots,\ \psi_n(t))$$

le vecteur fonction qui correspond à la trajectoire $x(t)$ en raison du principe du maximum.

On peut montrer que le vecteur $\psi(t_0)$ est *orthogonal à l'hypersurface S au point $x(t_0)$*, soit

$$\psi(t_0) = \lambda\ \text{grad}\ \omega(x(t_0)) \tag{21.7}$$

ou, ce qui revient au même,

$$\psi_i(t_0) = \lambda \frac{\partial\omega(x(t_0))}{\partial x_i} \qquad (i = 1,\ \ldots,\ n). \tag{21.8}$$

Conformément à (17.21) et (20.3), la fonction H est ici de la forme

$$H(\psi,\ x,\ v(x)) = \sum_{i=1}^{n} \psi_i f_i(x,\ v(x)). \tag{21.9}$$

D'après (9), (8) et (4)

$$H(\psi(t_0),\ x(t_0),\ v(x(t_0))) = \sum_{i=1}^{n} \psi_i(t_0) f_i(x(t_0),\ v(x(t_0))) =$$

$$= \lambda \sum_{i=1}^{n} \frac{\partial\omega(x(t_0))}{\partial x_i} f_i(x(t_0),\ v(x(t_0))) = \lambda. \tag{21.10}$$

Puisque (17.27) conduit à

$$H(\psi(t),\ x(t),\ v(x(t))) \geqslant 0,$$

on obtient de (10) et du point 2° du théorème 2 (p. 245) que $\lambda \geqslant 0$. (7) entraîne que $\lambda \neq 0$, puisque autrement on aurait $\psi(t_0) = 0$. Ainsi

$$\lambda > 0. \tag{21.11}$$

On déduit du principe du maximum que pour *tout* $u \in \Omega$

$$H(\psi(t_0),\ x(t_0),\ v(x(t_0))) \geqslant H(\psi(t_0),\ x(t_0),\ u). \qquad (21.12)$$

On trouve de (12), (4), (8), (10) et (11) que pour tout $u \in \Omega$

$$1 = \sum_{i=1}^{n} \frac{\partial \omega(x(t_0))}{\partial x_i} f_i(x(t_0),\ v(x(t_0))) = \frac{1}{\lambda} \sum_{i=1}^{n} \psi_i(t_0) f_i(x(t_0),\ v(x(t_0))) =$$

$$= \frac{1}{\lambda} H(\psi(t_0),\ x(t_0),\ v(x(t_0))) \geqslant \frac{1}{\lambda} H(\psi(t_0),\ x(t_0),\ u) =$$

$$= \frac{1}{\lambda} \sum_{i=1}^{n} \psi_i(t_0) f_i(x(t_0),\ u) = \sum_{i=1}^{n} \frac{\partial \omega(x(t_0))}{\partial x_i} f_i(x(t_0),\ u). \qquad (21.13)$$

(13) entraîne la relation

$$\max_{u \in \Omega} \sum_{i=1}^{n} \frac{\partial \omega(x(t_0))}{\partial x_i} f_i(x(t_0),\ u) = 1, \qquad (21.14)$$

qui coïncide avec l'équation de Bellman (15.7) du fait qu'on peut adopter comme point initial $x(t_0)$ tout point x (appartenant à l'ensemble $D \setminus M$) de la trajectoire marquée.

Ainsi, *si la condition* (7) *est respectée*, la fonction $\omega(x)$ satisfait sur l'ensemble $D \setminus M$ à l'équation de Bellman (15.7).

Pour achever la démonstration, il faut encore montrer que la relation (7) est vraie

$$\psi(t_0) = \lambda \operatorname{grad} \omega(x(t_0)).$$

La trajectoire marquée $x(t)$ est issue à l'instant t_0 du point $x(t_0)$ qui se trouve à l'intérieur de la case de première espèce σ_1.

Désignons par $\hat{x}(t)$ la trajectoire du système (20.3)

$$\frac{dx}{dt} = f(x,\ v(x)),$$

issue d'un point intérieur $\hat{x}(t_0)$ de cette même case σ_1. Considérons que le point $\hat{x}(t_0)$ se trouve suffisamment près du point $x(t_0)$ et, de plus, repose sur l'hypersurface S. Cette circonstance entraîne que les deux trajectoires $x(t)$ et $\hat{x}(t)$ issues à l'instant t_0 des points respectifs $x(t_0)$ et $\hat{x}(t_0)$ atteignent le point a au même instant $t_1 = -\omega(x(t_0)) + t_0 = -\omega(\hat{x}(t_0)) + t_0$.

Introduisons la notation ε pour la distance entre les points $x(t_0)$ et $\hat{x}(t_0)$ et considérons que ε est suffisamment petite.

Comme les points $x(t_0)$ et $\hat{x}(t_0)$ se trouvent dans la même case σ_1, les trajectoires $x(t)$ et $\hat{x}(t)$ rejoignent le point a en passant par la même suite des cases $\sigma_1, \sigma_2, \ldots, \sigma_q$, où σ_q est la case à une dimension adhérente au point a.

Nous avons montré plus haut que les instants τ_i $(i = 1, 2, \ldots, q - 1)$ de passage de la trajectoire marquée $x(t)$ d'une case

dans l'autre et les points correspondants de la trajectoire $x(\tau_i)$ sont des fonctions *dérivables* du point initial $x(t_0) \in \sigma_1$.

Désignons par δ_i l'intervalle de temps

$$\delta_i = \tau_i - \hat{\tau}_i \qquad (i = 1, 2, \ldots, q-1),$$

en posant $\tau_0 = t_0$, $\tau_q = t_1$.

Le fait que τ_i et $x(\tau_i)$ sont des fonctions dérivables du point $x(t_0) \in \sigma_1$ (et $\hat{\tau}_1$ et $\hat{x}(\hat{\tau}_i)$ sont des fonctions dérivables du point $\hat{x}(t_0) \in \sigma_1$) entraîne qu'il existe une constante C positive telle qu'aucun des intervalles de temps δ_i $(i = 1, 2, \ldots, q-1)$ ne dépasse en module $C\varepsilon$, et les trajectoires $x(t)$ et $\hat{x}(t)$ se trouvent l'une de l'autre à une distance de même ordre que ε:

$$|x(t) - \hat{x}(t)| \leqslant C\varepsilon \qquad (t_0 \leqslant t \leqslant t_1). \tag{21.15}$$

Effectuons maintenant quelques transformations. De même que précédemment, désignons par $\psi(t) = (\psi_1(t), \ldots, \psi_n(t))$ le vecteur fonction qui en raison du principe du maximum correspond à la trajectoire $x(t)$.

Puisque $x(t_1) = \hat{x}(t_1) = a$, on a

$$-\sum_{i=1}^{n} [x_i(t_0) - \hat{x}_i(t_0)]\,\psi_i(t_0) =$$

$$= -\sum_{i=1}^{n} [x_i(t_0) - \hat{x}_i(t_0)]\,\psi_i(t_0) + \sum_{i=1}^{n} [x_i(t_1) - \hat{x}_i(t_1)]\,\psi_i(t_1) =$$

$$= \int_{t_0}^{t_1} \frac{d}{dt}\left\{\sum_{i=1}^{n} [x_i(t) - \hat{x}_i(t)]\,\psi_i(t)\right\} dt =$$

$$= \int_{t_0}^{t_1} \left\{\sum_{i=1}^{n} \psi_i(t)\frac{d}{dt}[x_i(t) - \hat{x}_i(t)] + \sum_{i=1}^{n} [x_i(t) - \hat{x}_i(t)]\frac{d\psi_i(t)}{dt}\right\} dt =$$

$$= \int_{t_0}^{t_1} \left\{\sum_{i=1}^{n} \psi_i(t)\,[f_i(x(t), v(x(t))) - f_i(\hat{x}(t), v(\hat{x}(t)))] -\right.$$

$$\left. - \sum_{i=1}^{n} [x_i(t) - \hat{x}_i(t)]\,\frac{\partial H(\psi(t), x(t), v(x(t)))}{\partial x_i}\right\} dt =$$

$$= \int_{t_0}^{t_1} \left\{H(\psi(t), x(t), v(x(t))) - H(\psi(t), \hat{x}(t), v(\hat{x}(t))) -\right.$$

$$\left. - \sum_{i=1}^{n} [x_i(t) - \hat{x}_i(t)]\,\frac{\partial H(\psi(t), x(t), v(x(t)))}{\partial x_i}\right\} dt.$$

Ainsi

$$-\sum_{i=1}^{n}[x_i(t_0)-\hat{x}_i(t_0)]\,\psi_i(t_0)=\int_{t_0}^{t_1}F(t)\,dt, \qquad (21.16)$$

où

$$F(t)=H(\psi(t),\ x(t),\ v(x(t)))-H(\psi(t),\ \hat{x}(t),\ v(\hat{x}(t)))-$$
$$-\sum_{i=1}^{n}[x_i(t)-\hat{x}_i(t)]\,\frac{\partial H(\psi(t),\ x(t),\ v(x(t)))}{\partial x_i}. \qquad (21.17)$$

Si l'on désigne par ξ un certain point du segment qui relie les points $x(t)$ et $\hat{x}(t)$, en vertu du théorème connu de Lagrange on a la relation

$$H(\psi(t),\ \hat{x}(t),\ v(\hat{x}(t)))-H(\psi(t),\ x(t),\ v(\hat{x}(t)))=$$
$$=\sum_{i=1}^{n}[\hat{x}_i(t)-x_i(t)]\,\frac{\partial H(\psi(t),\ \xi,\ v(\hat{x}(t)))}{\partial x_i}. \qquad (21.18)$$

C'est pourquoi l'expression (17) peut s'écrire

$$F(t)=-H(\psi(t),x(t),v(\hat{x}(t)))+H(\psi(t),\ x(t),\ v(x(t)))-$$
$$-\sum_{i=1}^{n}[x_i(t)-\hat{x}_i(t)]\left[\frac{\partial H(\psi(t),x(t),v(x(t)))}{\partial x_i}-\frac{\partial H(\psi(t),\xi,v(\hat{x}(t)))}{\partial x_i}\right].$$
$$(21.19)$$

Comme $x(t)$ et $v(x(t))$ représentent la trajectoire marquée et la commande associée, qui d'après la condition E de synthèse régulière vérifient le principe du maximum, on a en vertu de la relation (17.28)

$$H(\psi(t),\ x(t),\ v(x(t)))\geqslant H(\psi(t),\ x(t),\ v(\hat{x}(t))). \qquad (21.20)$$

Il suit des relations (19) et (20) que

$$F(t)\geqslant G(t), \qquad (21.21)$$

où

$$G(t)=-\sum_{i=1}^{n}[x_i(t)-\hat{x}_i(t)]\left[\frac{\partial H(\psi(t),x(t),v(x(t)))}{\partial x_i}-\right.$$
$$\left.-\frac{\partial H(\psi(t),\xi,v(\hat{x}(t)))}{\partial x_i}\right]. \qquad (21.22)$$

Examinons maintenant les deux cas suivants. Dans le *premier* cas, le point t appartient à l'un des intervalles de temps Δ_i pendant lequel $x(t)$ et $\hat{x}(t)$ se trouvent dans la même case σ_i dans laquelle la fonction $v(x)$ est continûment dérivable. Compte tenu de (15), on a

l'estimation

$$|v(\hat{x}(t)) - v(x(t))| \leqslant C' \, |\hat{x}(t) - x(t)| \leqslant CC'\varepsilon. \quad (21.23)$$

Puisque le point ξ repose sur le segment qui relie les points $x(t)$ et $\hat{x}(t)$, on a d'après (15),

$$|\xi - x(t)| \leqslant |\hat{x}(t) - x(t)| \leqslant C\varepsilon. \quad (21.24)$$

La fonction $\frac{\partial H}{\partial x_i}$ étant continue par rapport à ses variables, pour $t \in \Delta_i$ la différence

$$\frac{\partial H(\psi(t), x(t), v(x(t)))}{\partial x_i} - \frac{\partial H(\psi(t), \xi, v(\hat{x}(t)))}{\partial x_i} \quad (21.25)$$

est un infiniment petit de même ordre que ε et tend avec ε vers zéro.

Donc, d'après (15) et (22), on a

$$\lim_{\varepsilon \to 0} \frac{G(t)}{\varepsilon} = 0 \text{ (uniformément par rapport à } t \in \Delta_i).$$

On en tire

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t_{k_i}}^{t_{k_i}+\Delta_i} G(t)\, dt = 0 \qquad (i = 1, 2, \ldots, q), \quad (21.26)$$

où t_{k_i} est l'origine de l'intervalle Δ_i.

Dans le *deuxième* cas, le point t appartient à l'un des intervalles de temps δ_i pendant lequel $x(t)$ et $\hat{x}(t)$ se trouvent dans deux cases voisines σ_i et σ_{i+1}. Ici on ne peut déjà plus affirmer que la différence (25) est infiniment petite avec ε, du fait que les points $x(t)$ et $\hat{x}(t)$ appartiennent à des cases différentes et lorsqu'on passe d'une case à l'autre la commande $v(x)$ peut subir des discontinuités.

Toutefois, pour tout t, la différence (25) est encore *bornée* du fait que la trajectoire $x(t)$ est bornée. C'est pourquoi, conformément à (15),

$$\lim_{\varepsilon \to 0} G(t) = 0 \text{ (uniformément par rapport à } t \in \delta_i).$$

La longueur du segment δ_i ne dépassant pas $C\varepsilon$ (cf. p. 284), il vient

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t_{k_i}+\Delta_i}^{t_{k_i}+\Delta_i+\delta_i} G(t)\, dt = 0 \qquad (i = 1, 2, \ldots, q-1). \quad (21.27)$$

En additionnant les relations (26) et (27), on obtient

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t_0}^{t_1} G(t)\, dt = 0. \quad (21.28)$$

Supposons maintenant que le point $\hat{x}(t_0)$ s'approche du point $x(t_0)$ suivant la surface S dans la direction d'un certain vecteur $p = (p_1, p_2, \ldots, p_n)$. On a alors

$$\lim_{\varepsilon\to 0} \frac{\hat{x}(t_0) - x(t_0)}{\varepsilon} = p. \tag{21.29}$$

D'après (16), (21) et (28), on a

$$\sum_{i=1}^{n} p_i \psi_i(t_0) = \lim_{\varepsilon\to 0} \sum_{i=1}^{n} \frac{1}{\varepsilon} [\hat{x}_i(t_0) - x_i(t_0)] \psi_i(t_0) =$$

$$= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{t_0}^{t_1} F(t)\, dt \geqslant \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{t_0}^{t_1} G(t)\, dt = 0. \tag{21.30}$$

Lorsque $\hat{x}(t_0) \to x(t_0)$ la trajectoire $\hat{x}(t) \to x(t)$ et la différence

$$F(t) - G(t) = H(\psi(t), x(t), v(x(t))) - H(\psi(t), x(t), v(\hat{x}(t))) \tag{21.31}$$

conditionnée par l'inégalité (20) peut être représentée par une expression analogue à (18). En reprenant les raisonnements employés pour déduire la relation (28), on trouve que

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{t_0}^{t_1} [F(t) - G(t)]\, dt = 0, \tag{21.32}$$

d'où, en vertu de (30), on obtient que pour tout vecteur p *tangent* à l'hypersurface S au point $x(t_0)$ on a la relation

$$\sum_{i=1}^{n} p_i \psi_i(t_0) = 0. \tag{21.33}$$

L'expression (33) entraîne la relation (7) ; le théorème est ainsi démontré.

§ 22. Liaison du principe du maximum avec la programmation dynamique

Considérons le problème à extrémité fixée de la trajectoire et à temps libre. Examinons le système décrit par les équations différentielles

$$\frac{dx_j}{dt} = f_j(x_1, \ldots, x_n, u_1, \ldots, u_r) \qquad (j = 1, \ldots, n), \tag{22.1}$$

équivalentes à l'équation vectorielle

$$\frac{dx}{dt} = f(x, u), \tag{22.2}$$

où x, u et f sont les vecteurs

$$x=\begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad u=\begin{bmatrix} u_1 \\ \cdots \\ u_r \end{bmatrix}, \quad f(x, u)=\begin{bmatrix} f_1(x, u) \\ \cdots\cdots \\ f_n(x, u) \end{bmatrix}. \tag{22.3}$$

On demande de ramener le système du point $x(t_0)$ de l'espace de phase X au point donné x^*. L'instant t_1 où le point représentatif tombe en x^* n'est pas fixé à l'avance.

La commande $u = u(t)$ doit satisfaire aux contraintes

$$u \in \Omega, \tag{22.4}$$

et il faut la choisir de façon que la fonctionnelle

$$Q=\int_{t_0}^{t_1} f_0(x(\xi), \ u(\xi))\, d\xi \tag{22.5}$$

soit minimisée.

Admettons que la commande qui satisfait à ces contraintes, la trajectoire qui lui correspond et l'intervalle de temps $t_1 - t_0$ sont optimaux.

La valeur minimisée $\Phi(x(t_0))$ de la fonctionnelle Q, obtenue dans ces conditions, est fonction de l'état initial $x(t_0)$ du système

$$\Phi(x(t_0)) = \Phi(x_1(t_0), x_2(t_0), \ldots, x_n(t_0)). \tag{22.6}$$

Nous avons montré au § 14 que sous l'hypothèse que la fonction $\Phi(x)$ est continue et possède partout des dérivées partielles continues par rapport à toutes ses variables, cette fonction vérifie l'équation non linéaire du premier ordre aux dérivées partielles (équation de Bellman) (14.36):

$$\min_{u \in \Omega} \left[f_0(x, u) + \sum_{i=1}^{n} \frac{\partial \Phi(x)}{\partial x_i} f_i(x, u) \right] = 0.$$

Introduisons, comme dans ce qui précède, la notation $x_0(t)$ pour la fonction scalaire déterminée par l'équation différentielle

$$\frac{dx_0}{dt} = f_0(x_1, \ldots, x_n, u_1, \ldots, u_r) = f_0(x, u) \tag{22.7}$$

et par la condition initiale

$$x_0(t_0) = 0. \tag{22.8}$$

On déduit de (7), (8) et (5) que la fonctionnelle Q à minimiser peut s'écrire

$$Q = x_0(t_1). \tag{22.9}$$

Introduisons les vecteurs de dimension $(n+1)$:

$$\tilde{x}=\begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix}, \quad \tilde{f}=\begin{bmatrix} f_0 \\ f_1 \\ \dots \\ f_n \end{bmatrix}. \tag{22.10}$$

Désignons par $\tilde{\Phi}(\tilde{x})$ la fonction

$$\tilde{\Phi}(\tilde{x}) = x_0 + \Phi(x_1, x_2, \dots, x_n) = x_0 + \Phi(x). \tag{22.11}$$

L'équation de Bellman (14.36) pourra maintenant s'écrire

$$\min_{u\in\Omega} \sum_{i=0}^{n} \frac{\partial\tilde{\Phi}(\tilde{x})}{\partial x_i} f_i(x, u) = 0. \tag{22.12}$$

Si l'on multiplie les deux membres de (12) par -1, le signe min du premier membre de (12) doit être remplacé par max. Ainsi, l'équation (12) peut se mettre sous la forme

$$\max_{u\in\Omega} \sum_{i=0}^{n} \left[-\frac{\partial\tilde{\Phi}(\tilde{x})}{\partial x_i}\right] f_i(x, u) = 0. \tag{22.13}$$

D'une façon analogue à (15.28), introduisons les notations

$$\omega(x) = -\Phi(x), \tag{22.14}$$
$$\tilde{\omega}(\tilde{x}) = -\tilde{\Phi}(\tilde{x}). \tag{22.15}$$

L'équation de Bellman (13) devient alors

$$\max_{u\in\Omega} \sum_{i=0}^{n} \frac{\partial\tilde{\omega}(\tilde{x})}{\partial x_i} f_i(x, u) = 0. \tag{22.16}$$

Dans ces conditions la borne supérieure est atteinte pour une certaine $u \in \Omega$, ou, plus précisément, pour la valeur que prend la commande optimale à l'instant de la sortie du point x.

Enonçons encore les hypothèses supplémentaires suivantes. Supposons que la fonction $\omega(x)$ possède des dérivées secondes continues $\frac{\partial^2\omega(x)}{\partial x_i \partial x_j}$ $(i, j=1, \dots, n)$, et les fonctions $f_i(x, u)$ possèdent des dérivées continues

$$\frac{\partial f_i(x, u)}{\partial x_j} \qquad (i=0, 1, \dots, n;\ j=1, \dots, n).$$

La fonction sous le signe max du premier membre de (16) se note

$$g(\tilde{x}, u) = \sum_{i=0}^{n} \frac{\partial\tilde{\omega}(\tilde{x})}{\partial x_i} f_i(x, u). \tag{22.17}$$

Soit $u(t)$ $(t_0 \leqslant t \leqslant t_1)$ la *commande optimale* qui ramène le point représentatif de l'état $x(t_0)$ à l'état x^*, et $x(t)$, la trajectoire optima-

le associée. Fixons un certain instant t ($t_0 \leqslant t \leqslant t_1$) et considérons la fonction $g(\tilde{x}, u(t))$ de la variable $\tilde{x}$. Conformément aux hypothèses précédentes, la fonction $g(\tilde{x}, u(t))$ possède des dérivées continues par rapport aux variables x_i ($i = 0, 1, \ldots, n$):

$$\frac{\partial g(\tilde{x}, u(t))}{\partial x_k} = \sum_{i=0}^{n} \frac{\partial^2 \tilde{\omega}(\tilde{x})}{\partial x_i \partial x_k} f_i(x, u(t)) + \sum_{i=0}^{n} \frac{\partial \tilde{\omega}(\tilde{x})}{\partial x_i} \frac{\partial f_i(x, u(t))}{\partial x_k} \tag{22.18}$$

$$(k = 0, 1, \ldots, n).$$

L'équation (16) entraîne que tout mouvement optimal $(x(t), u(t))$ donne lieu à la relation

$$g(\tilde{x}(t), u(t)) \equiv 0. \tag{22.19}$$

Les points x qui à l'instant fixé ne reposent pas sur la trajectoire optimale, appartiennent à d'autres trajectoires réalisées par des commandes admissibles mais non optimales. C'est pourquoi en raison de l'équation (16), en ces points

$$g(\tilde{x}, u) \leqslant 0. \tag{22.20}$$

Par conséquent, la fonction $g(\tilde{x}, u(t))$, éxaminée au (18), de la variable x atteint au point $\tilde{x} = \tilde{x}(t)$ le maximum (l'instant t est fixé); donc, en ce point, ses dérivées partielles par rapport à x_0, $x_1, \ldots, x_n$ s'annulent. D'après (18), on obtient ainsi les relations suivantes observées le long de la trajectoire optimale

$$\sum_{i=0}^{n} \frac{\partial^2 \tilde{\omega}(\tilde{x}(t))}{\partial x_i \partial x_k} f_i(x(t), u(t)) + \sum_{i=0}^{n} \frac{\partial \tilde{\omega}(\tilde{x}(t))}{\partial x_i} \frac{\partial f_i(x(t), u(t))}{\partial x_k} = 0 \tag{22.21}$$

$$(k = 0, 1, \ldots, n;\ t_0 \leqslant t \leqslant t_1).$$

En tenant compte que d'après (1) et (7) le mouvement optimal $(x(t), u(t))$ donne lieu à la relation

$$\frac{dx_i(t)}{dt} = f_i(x(t), u(t)) \qquad (i = 0, 1, \ldots, n), \tag{22.22}$$

on a

$$\frac{d}{dt}\left(\frac{\partial \tilde{\omega}(\tilde{x}(t))}{\partial x_k}\right) = \sum_{i=0}^{n} \frac{\partial^2 \tilde{\omega}(\tilde{x}(t))}{\partial x_i \partial x_k} \frac{dx_i(t)}{dt} =$$

$$= \sum_{i=1}^{n} \frac{\partial^2 \tilde{\omega}(\tilde{x}(t))}{\partial x_i \partial x_k} f_i(x(t), u(t)). \tag{22.23}$$

En vertu de (23), la relation (21) s'écrit

$$\frac{d}{dt}\left(\frac{\partial\tilde{\omega}(\tilde{x}(t))}{\partial x_k}\right)+\sum_{i=0}^{n}\frac{\partial\tilde{\omega}(\tilde{x}(t))}{\partial x_i}\,\frac{\partial f_i(x(t),u(t))}{\partial x_k}=0 \qquad (22.24)$$

$$(k=0,\ 1,\ \ldots,\ n).$$

Désignons par $\psi_k(t)$ $(k = 0, 1, \ldots, n)$ la fonction

$$\psi_k(t)=\frac{\partial\tilde{\omega}(\tilde{x}(t))}{\partial x_k} \qquad (k=0,\ 1,\ \ldots,\ n). \qquad (22.25)$$

Il résulte de (24) que le long de toute trajectoire optimale les grandeurs $\psi_k(t)$ satisfont au système suivant d'équations différentielles linéaires

$$\frac{d\psi_k(t)}{dt}=-\sum_{i=0}^{n}\frac{\partial f_i(x(t),u(t))}{\partial x_k}\psi_i(t) \qquad (k=0,\ 1,\ \ldots,\ n). \qquad (22.26)$$

Remarquons que d'après (25), (15) et (11),

$$\psi_0(t) = -1, \qquad (22.27)$$

ce qui ne contredit pas les équations (26) du fait que d'après (1) et (5), $f_i(x(t), u(t))$ $(i = 0, 1, \ldots, n)$ ne dépendent pas de x_0 et pour $k = 0$ l'équation (26) se met sous la forme

$$\frac{d\psi_0(t)}{dt}=0. \qquad (22.28)$$

Le vecteur ψ dont les coordonnées ψ_k $(k = 0, 1, \ldots, n)$ sont déterminées par les relations (25) a pour expression

$$\psi=\begin{bmatrix}\psi_0\\ \psi_1\\ \cdots\\ \psi_n\end{bmatrix}=\begin{bmatrix}\dfrac{\partial\tilde{\omega}}{\partial x_0}\\ \dfrac{\partial\tilde{\omega}}{\partial x_1}\\ \cdots\\ \dfrac{\partial\tilde{\omega}}{\partial x_n}\end{bmatrix}=\begin{bmatrix}-1\\ \dfrac{\partial\tilde{\omega}}{\partial x_1}\\ \cdots\\ \dfrac{\partial\tilde{\omega}}{\partial x_n}\end{bmatrix}. \qquad (22.29)$$

Conformément à (25), (10) et (29), l'équation de Bellman (16) devient

$$\sum_{i=0}^{n}\psi_i(t)\,f_i(x(t),\ u(t))=\max_{u\in\Omega}\sum_{i=0}^{n}\psi_i(t)\,f_i(x(t),\ u)=$$
$$=\max_{u\in\Omega}\langle\psi(t),\ \tilde{f}(x(t),\ u)\rangle=0. \qquad (22.30)$$

Désignons par $\tilde{H}(\psi, x, u)$ la fonction

$$\tilde{H}(\psi,\ x,\ u)=\sum_{i=0}^{n}\psi_i f_i(x,\ u)=\langle\psi,\ \tilde{f}(x,\ u)\rangle, \qquad (22.31)$$

qui dépend de $2n + r + 1$ variables $\psi_0, \psi_1, \ldots, \psi_n, x_1, \ldots, x_n, u_1, \ldots, u_r$.

L'équation (30) se met sous la forme

$$\max_{u\in\Omega} \widetilde{H}(\psi, x, u) = 0. \tag{22.32}$$

La relation (32) permet de conclure qu'à tout instant t la commande optimale $u(t)$ maximise la fonction $\widetilde{H}$ par rapport à toute autre commande $u \in \Omega$ qui ramène le système du point $x(t_0)$ au point x^*. Cette valeur maximale est la même, plus précisément elle est nulle en tout point de la trajectoire optimale, c'est-à-dire l'égalité

$$\widetilde{H}(\psi(t), x(t), u(t)) \equiv 0 \tag{22.33}$$

est vérifiée le long de la trajectoire optimale.

Notons, toutefois, que les relations (30), (32) et (33) sont obtenues ici sous l'hypothèse initiale que la fonction $\Phi(x_1, \ldots, x_n)$ admet des dérivées partielles continues par rapport à toutes ses variables, ce qui, comme nous l'avons montré plus haut (§ 20), n'est pas le fait même de problèmes relativement simples. C'est pourquoi les transformations données dans le présent paragraphe et les corrolaires qui en découlent n'ont qu'un caractère euristique et ne peuvent servir que d'indications suggestives.

§ 23. Exemples d'application du principe du maximum

1. Théorème sur le nombre de commutations d'une commande dans le problème linéaire de la réponse en temps minimal. Considérons un système régi par l'équation différentielle vectorielle linéaire

$$\frac{dx}{dt} = Ax + Bu, \tag{23.1}$$

où x, A, B sont les matrices

$$x = \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & A_{12} \ldots A_{1n} \\ \cdots & \cdots \\ A_{n1} & A_{n2} \ldots A_{nn} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ \cdots \\ B_n \end{bmatrix}, \tag{23.2}$$

et où la commande u est une fonction scalaire soumise aux contraintes

$$-1 \leqslant u \leqslant 1. \tag{23.3}$$

On demande de faire passer le système du point $x(0)$, où le système se trouve à l'instant t_0, au point $x = 0$ en un temps minimal.

Désignons par ψ le vecteur de dimension n

$$\psi = \begin{bmatrix} \psi_1 \\ \cdots \\ \psi_n \end{bmatrix}, \tag{23.4}$$

et introduisons d'après (17.21) la fonction

$$H(\psi, x, u) = \sum_{j=1}^{n} \sum_{k=1}^{n} \psi_j A_{jk} x_k + \sum_{j=1}^{n} B_j \psi_j u = \langle \psi, Ax \rangle + \langle \psi, Bu \rangle. \quad (23.5)$$

D'après (17.23), les fonctions ψ_j $(j = 1, \ldots, n)$ vérifient les équations différentielles

$$\frac{d\psi_j}{dt} = -\sum_{k=1}^{n} A_{kj} \psi_k \qquad (j = 1, \ldots, n). \quad (23.6)$$

Le système d'équations différentielles scalaires (6) est équivalent à l'équation différentielle vectorielle

$$\frac{d\psi}{dt} = -A^* \psi, \quad (23.7)$$

où, d'après (6)

$$A^* = \begin{bmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{bmatrix}. \quad (23.8)$$

La matrice A^* est la transposée de la matrice A.

L'équation différentielle vectorielle (7) est conjuguée à l'équation

$$\frac{d\xi}{dt} = A\xi. \quad (23.9)$$

Comme nous l'avons montré plus haut (18.51), les intégrales des équations (9) et (7) jouissent de la propriété suivante:

$$\langle \xi(t), \psi(t) \rangle = \langle \xi(0), \psi(0) \rangle.$$

D'après (18.53), la matrice fondamentale $X(t)$ des solutions de l'équation (9) et la matrice fondamentale $\Xi(t)$ des solutions de l'équation (7) vérifient la relation

$$\Xi^*(t) = X^{-1}(t).$$

D'après (10.160)

$$X^{-1}(t) = X(-t),$$

et donc

$$\Xi^*(t) = X(-t), \quad (23.10)$$

d'où il résulte que les racines des équations caractéristiques $\det(\lambda E - A) = 0$ et $\det(\mu E + A^*) = 0$ associées aux équations différentielles (9) et (7) *ne diffèrent que par les signes* ([28], p. 171).

En vertu du théorème 2 (§ 17), la commande optimale est

$$u(t) = \operatorname{sign} \sum_{j=1}^{n} B_j \psi_j(t). \quad (23.11)$$

Considérons le cas où toutes les racines λ_i ($i = 1, \ldots, n$) de l'équation caractéristique

$$\det (\lambda E - A) = 0$$

sont simples et représentent des grandeurs réelles.

Dans ce cas, la propriété du système conjugué indiquée ci-dessus fait que la solution de l'équation différentielle vectorielle (7) est la suivante :

$$\psi_j(t) = \sum_{l=1}^{n} C_{jl} e^{\mu_l t} \qquad (j = 1, \ldots, n), \tag{23.12}$$

où μ_l ($l = 1, \ldots, n$) sont des variables réelles, et C_{jl} ($j, l = 1, \ldots, n$) certaines constantes qui dépendent de n constantes arbitraires qu'on détermine en donnant les valeurs initiales $\psi_j(0)$ ($j = 1, \ldots, n$) des fonctions $\psi_j(t)$ (ou les valeurs que doivent prendre les fonctions $\psi_j(t)$ à tout autre instant fixé).

En portant dans (11) les expressions (12), on obtient

$$u(t) = \operatorname{sign}(S_1 e^{\mu_1 t} + S_2 e^{\mu_2 t} + \ldots + S_n e^{\mu_n t}), \tag{23.13}$$

où S_j sont des constantes

$$S_j = \sum_{\nu=1}^{n} B_\nu C_{\nu j} \qquad (j = 1, \ldots, n). \tag{23.14}$$

Puisque le nombre de zéros de la fonction

$$s(t) = \sum_{j=1}^{n} S_j e^{\mu_j t} \tag{23.15}$$

dans l'intervalle $0 < t < \infty$ *ne dépasse pas* $n - 1$, la fonction $u(t)$ définie par l'expression (11) et dont le module $|u(t)| = 1$ est une fonction constante par morceaux possède dans l'intervalle de temps $0 < t < t_1$ au plus $n - 1$ points de discontinuité, c'est-à-dire la fonction $u(t)$ possède au plus n intervalles de permanence.

Autrement dit, le nombre de commutations de la commande $u(t)$ *ne dépasse pas* $n - 1$.

Le théorème démontré s'appelle théorème de n intervalles; il est dû à A. Feldbaum [85].

2. Problème de l'écart maximal. Considérons le système décrit par l'équation différentielle scalaire

$$\frac{dx_1}{dt} = f_1(x_1, u), \tag{23.16}$$

où

$$f_1(x_1, u) = -ax_1 + bu. \tag{23.17}$$

La commande u subit la contrainte

$$-1 \leqslant u \leqslant 1. \tag{23.18}$$

On demande de choisir la commande u telle que l'écart du système $x_1(T)$ à l'instant T soit le plus grand possible.

Bien entendu pour un système décrit par une équation aussi simple que (16), la solution du problème peut être immédiate.

Toutefois, cherchons sa solution à l'aide du principe du maximum pour illustrer la méthode générale par un exemple exempt de calculs trop longs.

Désignons par Q la fonctionnelle

$$Q = \int_0^T f_0(x_1, u)\, dt, \tag{23.19}$$

avec

$$f_0(x_1, u) = -f_1(x_1, u) = ax_1 - bu. \tag{23.20}$$

Introduisons la notation $x_0(t)$ pour la fonction scalaire qui vérifie l'équation différentielle

$$\frac{dx_0}{dt} = f_0(x_1, u) \tag{23.21}$$

et la condition initiale

$$x_0(0) = 0. \tag{23.22}$$

On tire de (19), (21) et (22) que

$$Q = x_0(T). \tag{23.23}$$

D'après (16), (21) et (20),

$$\frac{dx_0}{dt} = -\frac{dx_1}{dt} \tag{23.24}$$

d'où, en tenant compte de (22), on a

$$x_0(T) = -x_1(T) + x_1(0). \tag{23.25}$$

L'écart initial $x_1(0)$ ne dépendant pas de la commande, la relation (25) entraîne que la commande $u(t)$, minimisant la fonctionnelle Q, maximise la valeur possible de l'écart $x_1(T)$.

Désignons par ψ le vecteur

$$\psi = \begin{bmatrix} \psi_0 \\ \psi_1 \end{bmatrix}, \tag{23.26}$$

et introduisons d'après (17.10) la fonction

$$\tilde{H}(\psi, x, u) = \psi_0 f_0(x_1, u) + \psi_1 f_1(x_1, u) =$$
$$= a(\psi_0 - \psi_1) x_1 + b(-\psi_0 + \psi_1) u. \tag{23.27}$$

D'après (17.14), les fonctions ψ_0 et ψ_1 vérifient les équations

$$\frac{d\psi_0}{dt} = -\frac{\partial \widetilde{H}}{\partial x_0}, \quad \frac{d\psi_1}{dt} = -\frac{\partial \widetilde{H}}{\partial x_1}, \tag{23.28}$$

qui, conformément à (27), se mettent sous la forme

$$\frac{d\psi_0}{dt} = 0, \quad \frac{d\psi_1}{dt} = a\psi_1 - a\psi_0. \tag{23.29}$$

La solution générale des équations (29) est

$$\psi_0 = C_1, \quad \psi_1(t) = C_2 e^{at} + C_1. \tag{23.30}$$

Le problème concerné est un problème à temps fixé T et à extrémité libre de la trajectoire. C'est pourquoi, d'après (17.55), les fonctions $\psi_0(t)$ et $\psi_1(t)$ doivent vérifier les conditions

$$\psi_0(T) = -1, \quad \psi_1(T) = 0. \tag{23.31}$$

Sous les conditions (31) les solutions (30) deviennent

$$\psi_0(t) = -1, \quad \psi_1(t) = e^{a(t-T)} - 1. \tag{23.32}$$

Si la commande s'écrit

$$u(t) = \operatorname{sign}[-\psi_0(t) + \psi_1(t)], \tag{23.33}$$

pour $b > 0$ la fonction $\widetilde{H}(\psi(t), x(t), u)$ définie par l'expression (27) de la variable u (soumise aux contraintes (18)) atteint le maximum au point $u = u(t)$ pour tout t $(0 \leqslant t \leqslant T)$.

En remplaçant dans (33) $\psi_0(t)$ et $\psi_1(t)$ par leurs expressions (32), on obtient

$$u(t) = \operatorname{sign} e^{a(t-T)} = 1. \tag{23.34}$$

Sous la condition (34) on trouve à partir de (16) et (17)

$$\max_{-1 \leqslant u \leqslant 1} x_1(T) = x_1(0)\, e^{-aT} + \frac{b}{a}(1 - e^{-aT}), \tag{23.35}$$

ce qui, comme nous l'avons déjà dit, est évident dans ce problème.

3. Application du principe du maximum à une commande non soumise aux contraintes. En l'absence des contraintes, dans de nombreux cas il est possible d'obtenir sous une forme explicite la solution du problème du choix de la commande minimisant une fonctionnelle.

Etudions le système décrit par l'équation différentielle scalaire (cf. § 16, point 2)

$$\frac{dq_1}{dt} = f_1(q_1, u), \tag{23.36}$$

où

$$f_1(q_1, u) = \frac{1}{m} u, \tag{23.37}$$

et dont l'état initial est

$$[q_1(t)]_{t=0} = q_1(0).$$

On demande d'amener le système à l'instant t_1 au point $q_1(t_1) = 0$ en choisissant la commande $u = u(t)$ minimisant la fonctionnelle

$$Q = \int_0^{t_1} f_0(q_1, u)\, dt, \tag{23.38}$$

avec

$$f_0(q_1, u) = -\frac{c}{2} q_1^2 + \frac{1}{2m} u^2. \tag{23.39}$$

Conformément à (17.10) construisons la fonction

$$\widetilde{H}(\psi, q_1, u) = \psi_0 f_0(q_1, u) + \psi_1 f_1(q_1, u) = \\ = \psi_0\left(-\frac{c}{2} q_1^2 + \frac{1}{2m} u^2\right) + \frac{1}{m}\psi_1 u \tag{23.40}$$

et composons les équations

$$\frac{dq_j}{dt} = \frac{\partial \widetilde{H}}{\partial \psi_j}, \quad \frac{d\psi_j}{dt} = -\frac{\partial \widetilde{H}}{\partial q_j} \qquad (j = 0,\ 1). \tag{23.41}$$

D'après (40), les équations (41) deviennent

$$\frac{dq_0}{dt} = -\frac{c}{2} q_1^2 + \frac{1}{2m} u^2, \quad \frac{dq_1}{dt} = \frac{1}{m} u, \tag{23.42}$$

$$\frac{d\psi_0}{dt} = 0, \quad \frac{d\psi_1}{dt} = cq_1\psi_0. \tag{23.43}$$

Cherchons maintenant la commande u (ne supportant aucune contrainte fixée à l'avance) maximisant la fonction $\widetilde{H}$. Etant donné que

$$\frac{\partial \widetilde{H}}{\partial u} = \frac{1}{m}(\psi_0 u + \psi_1), \quad \frac{\partial^2 \widetilde{H}}{\partial u^2} = \frac{1}{m}\psi_0, \tag{23.44}$$

sous la condition

$$\psi_0 < 0, \tag{23.45}$$

la fonction $\widetilde{H}$ possède un maximum par rapport à u avec

$$u = -\frac{\psi_1}{\psi_0}. \tag{23.46}$$

La première équation (43) entraîne

$$\psi_0 = \text{const.} \tag{23.47}$$

En portant dans la deuxième équation (42) la valeur optimale de la commande u définie par l'expression (46), on a

$$\frac{dq_1}{dt} = -\frac{1}{m}\frac{\psi_1}{\psi_0}. \tag{23.48}$$

En dérivant par rapport à t les deux membres de l'équation (48) et en tenant compte que $\psi_0 = \text{const}$, on obtient

$$\frac{d^2q_1}{dt^2} = -\frac{1}{m\psi_0}\frac{d\psi_1}{dt}. \tag{23.49}$$

En vertu de la deuxième équation (43), l'équation (49) s'écrit

$$\frac{d^2q_1}{dt^2} = -k^2q_1, \tag{23.50}$$

où

$$k^2 = \frac{c}{m}. \tag{23.51}$$

Nous avons ainsi abouti à un problème aux limites qui consiste à résoudre l'équation différentielle (50) vérifiant les conditions

$$[q_1(t)]_{t=0} = q_1(0), \quad [q_1(t)]_{t=t_1} = 0. \tag{23.52}$$

Voici la solution que nous recherchons

$$q_1(t) = -\frac{q_1(0)\cos kt_1}{\sin kt_1}\sin kt + q_1(0)\cos kt. \tag{23.53}$$

La solution (53) existe sous la condition

$$t_1 \neq \frac{n\pi}{k} \qquad (n = 1, 2, 3, \ldots), \tag{23.54}$$

que nous allons supposer remplie.

Il résulte de la deuxième équation (42) que

$$u = m\frac{dq_1}{dt},$$

d'où, en vertu de (53), on obtient

$$u = -\sqrt{mc}\,q_1(0)\left[\frac{\cos kt_1}{\sin kt_1}\cos kt + \sin kt\right]. \tag{23.55}$$

En comparant l'expression (46), obtenue d'après le principe du maximum de Pontriaguine, avec la deuxième équation de (42), on en tire que dans ce problème

$$-\frac{\psi_1}{\psi_0} = m\frac{dq_1}{dt}, \tag{23.56}$$

d'où

$$\psi_1(t) = \sqrt{mc}\,q_1(0)\left[\frac{\cos kt_1}{\sin kt_1}\cos kt + \sin kt\right]\psi_0 \tag{23.57}$$

Constatons que pour le système mécanique dont la fonction de Lagrange est

$$L = \frac{1}{2} m\dot{q}_1^2 - \frac{1}{2} cq_1^2, \tag{23.58}$$

l'impulsion canonique p_1 est

$$p_1 = \frac{\partial L}{\partial \dot{q}_1} = m\dot{q}_1. \tag{23.59}$$

Il en résulte que pour un système commandé décrit par l'équation (36)

$$u = p_1, \tag{23.60}$$

$$\psi_1 = -\psi_0 p_1, \tag{23.61}$$

où p_1 est l'impulsion canonique d'un système mécanique (58).

Les relations (60) et (61) ont lieu parce que dans ce cas, de même qu'au § 16, point 2, la fonctionnelle (38) est choisie telle que pour le système mécanique (58) elle devient

$$Q = \int_0^{t_1} L\,dt, \tag{23.62}$$

ce qui représente l'action hamiltonienne, qui d'après le principe de Hamilton, prend une valeur stationnaire pour des mouvements réels d'un système mécanique.

En substituant dans (40) à u et ψ_1 les expressions (60) et (61), on obtient pour la fonction $\tilde{H}$

$$\tilde{H} = -\psi_0 \left(\frac{p_1^2}{2m} + \frac{cq_1^2}{2}\right), \tag{23.63}$$

c'est-à-dire la fonction $\tilde{H}$ de notre exemple est le produit de la fonction de Hamilton du système mécanique (58) par la constante $-\psi_0$.

§ 24. Systèmes linéaires optimaux à critère quadratique de performance

1. Problème du régulateur d'état [34, 52]. Considérons le système linéaire non stationnaire

$$\dot{x} = A(t)\,x + B(t)\,u(t), \tag{24.1}$$

où x est le vecteur de dimension n des coordonnées de phase du système; $u(t)$, le vecteur de dimension r de la commande; $A(t)$, la matrice $n \times n$; $B(t)$, la matrice $n \times r$. Supposons que les commandes $u_j(t)$ $(j = 1, \ldots, r)$ ne soient pas soumises à des contraintes. La

commande $u(t)$ doit être choisie de façon à minimiser la fonctionnelle

$$J_1 = \frac{1}{2}\langle x(T), Fx(T)\rangle + \frac{1}{2}\int_{t_0}^{T} [\langle x(t), Q(t)\, x(t)\rangle + \langle u(t), R(t)\, u(t)\rangle]\, dt. \tag{24.2}$$

Ici T est un instant fixé. La commande a pour but de maintenir les coordonnées de phase du système $x_j(t)$ ($j = 1, \ldots, n$) dans le voisinage du zéro, c'est-à-dire il s'agit d'assurer la régulation de l'état du système. F et $Q(t)$ désignent des matrices $n \times n$ définies non négatives, c'est-à-dire des matrices symétriques vérifiant la condition $\langle a, Fa\rangle \geqslant 0$, $\langle a, Q(t)\, a\rangle \geqslant 0$ pour tout vecteur $a \neq 0$ de dimension n. La matrice $R(t)$ est une matrice définie positive $r \times r$, c'est-à-dire une matrice symétrique vérifiant la condition $\langle b, R(t)\, b\rangle > 0$ pour tout vecteur $b \neq 0$ de dimension r.

Les conditions nécessaires d'optimalité de la commande $u(t)$ peuvent s'obtenir à partir du principe du maximum de Pontriaguine. D'après (18.15), la fonction $\widetilde{H}$ du problème que nous étudions s'écrit

$$\widetilde{H} = \frac{1}{2}\psi_0(t)\langle x(t), Q(t)\, x(t)\rangle + \frac{1}{2}\psi_0(t)\langle u(t), R(t)\, u(t)\rangle + \\ + \langle A(t)\, x(t), \psi(t)\rangle + \langle B(t)\, u(t), \psi(t)\rangle, \tag{24.3}$$

où $\psi(t)$ est le vecteur de dimension n dont les éléments sont les variables auxiliaires $\psi_1(t), \ldots, \psi_n(t)$. La commande optimale $u(t)$ doit maximiser la fonction $\widetilde{H}$. D'après le principe du maximum appliqué aux systèmes non autonomes, dans le problème à temps fixé et à extrémité libre de la trajectoire $\psi_0(t) = \text{const} = -1$. C'est pourquoi en introduisant la notation

$$p_i(t) = \frac{\psi_i(t)}{\psi_0} \qquad (i = 1, \ldots, n), \tag{24.4}$$

on obtient que la commande optimale $u(t)$ minimise la fonction

$$\mathscr{H} = \frac{1}{2}\langle x(t), Q(t)\, x(t)\rangle + \frac{1}{2}\langle u(t), R(t)\, u(t)\rangle + \\ + \langle A(t)\, x(t), p(t)\rangle + \langle B(t)\, u(t), p(t)\rangle. \tag{24.5}$$

Puisque

$$\dot{\psi}_j = \psi_0 \dot{p}_j = -\frac{\partial \widetilde{H}}{\partial x_j} \qquad (j = 1, \ldots, n),$$

les éléments du vecteur p vérifient le système d'équations différentielles

$$\dot{p}_j = -\frac{\partial \mathscr{H}}{\partial x_j} \qquad (j = 1, \ldots, n). \tag{24.6}$$

Les dérivées partielles par rapport à x_j $(j = 1, \ldots, n)$ de la fonction $\Xi = \frac{1}{2} \langle x(t), Q(t)\, x(t) \rangle$, où $Q(t)$ est une matrice symétrique, s'écrivent

$$\frac{\partial \Xi}{\partial x_j} = \sum_{k=1}^{n} Q_{jk}(t)\, x_k(t) \qquad (j = 1, \ldots, n),$$

et le vecteur grad $\Xi = \left(\frac{\partial \Xi}{\partial x_1}, \frac{\partial \Xi}{\partial x_2}, \ldots, \frac{\partial \Xi}{\partial x_n}\right)$ est de la forme

$$\text{grad } \Xi = Q(t)\, x(t).$$

D'une façon analogue, pour la fonction $\mathcal{L} = \langle A(t)\, x(t), p(t) \rangle$ on a

$$\frac{\partial \mathcal{L}}{\partial x_j} = \sum_{k=1}^{n} A_{kj}(t)\, p_k(t) \qquad (j = 1, \ldots, n),$$

et le vecteur grad $\mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial x_1}, \frac{\partial \mathcal{L}}{\partial x_2}, \ldots, \frac{\partial \mathcal{L}}{\partial x_n}\right)$ s'écrit

$$\text{grad } \mathcal{L} = A^*(t)\, p(t),$$

où $A^*(t)$ désigne la matrice obtenue par transposition de $A(t)$.

Ainsi, conformément à (6) et à (5), le vecteur fonction $p(t)$ satisfait à l'équation différentielle

$$\dot{p} = -Q(t)\, x - A^*(t)\, p. \qquad (24.7)$$

Cherchons maintenant la commande $u(t)$ qui extrémise la fonction $\mathcal{H}$. Comme

$$\frac{\partial \mathcal{H}}{\partial u_j} = \sum_{l=1}^{r} R_{jl}(t)\, u_l(t) + \sum_{k=1}^{n} B_{kj}(t)\, p_k(t) \qquad (j = 1, \ldots, r), \qquad (24.8)$$

les conditions nécessaires pour l'obtention de l'extrémum

$$\frac{\partial \mathcal{H}}{\partial u_j} = 0 \qquad (j = 1, \ldots, r)$$

sont de la forme

$$\sum_{l=1}^{r} R_{jl}(t)\, u_l(t) = -\sum_{k=1}^{n} B_{kj}(t)\, p_k(t) \qquad (j = 1, \ldots, r). \qquad (24.9)$$

Le système d'équations scalaires (9) est équivalent à l'équation vectorielle

$$R(t)\, u(t) = -B^*(t)\, p(t). \qquad (24.10)$$

L'équation (10) entraîne que la commande $u(t)$ qui fournit l'extrémum à la fonction $\mathscr{H}$ est

$$u(t) = -R^{-1}(t)\, B^*(t)\, p(t). \qquad (24.11)$$

Puisque $R(t)$ est une matrice définie positive, $\det R(t) \neq 0$ et il existe la matrice inverse $R^{-1}(t)$.

D'après (8)

$$\frac{\partial^2 \mathscr{H}}{\partial u_j \partial u_k} = R_{jk}(t) \qquad (j, k = 1, \ldots, r),$$

et, par suite, la matrice

$$\begin{bmatrix} \frac{\partial^2 \mathscr{H}}{\partial u_1 \partial u_1} & \cdots & \frac{\partial^2 \mathscr{H}}{\partial u_1 \partial u_r} \\ \cdot & \cdot\,\cdot\,\cdot & \cdot \\ \frac{\partial^2 \mathscr{H}}{\partial u_r \partial u_1} & \cdots & \frac{\partial^2 \mathscr{H}}{\partial u_r \partial u_r} \end{bmatrix} = R(t) \qquad (24.12)$$

est définie positive. Il s'ensuit que la commande $u(t)$, définie par l'expression (11), minimise la fonction $\mathscr{H}$.

Portons l'expression (11) de $u(t)$ dans l'équation (1) et ramenons cette équation à la forme

$$\dot{x} = A(t)\, x - S(t)\, p, \qquad (24.13)$$

où $S(t)$ est une matrice symétrique $n \times n$:

$$S(t) = B(t)\, R^{-1}(t)\, B^*(t). \qquad (24.14)$$

Les équations (13) et (7) forment un système d'équations différentielles vectorielles qu'on peut écrire

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} A(t) & -S(t) \\ -Q(t) & -A^*(t) \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix}. \qquad (24.15)$$

L'équation différentielle vectorielle (15) décrit précisément le système optimal recherché.

L'état initial $x(t_0)$ du système (1) est supposé connu. Pour cette raison la solution du système (15) de rang $2n$ doit vérifier n conditions initiales définies par la relation vectorielle

$$[x(t)]_{t=t_0} = x(t_0). \qquad (24.16)$$

Pour les variables auxiliaires p_j ($j = 1, \ldots, n$), les n conditions aux limites qui résultent des conditions de transversalité sont définies par la relation

$$p(T) = Fx(T). \qquad (24.17)$$

Les conditions (16) et (17) déterminent la solution unique du système d'équations (15).

Montrons comment obtenir la condition aux limites (17). Ramenons au préalable la fonctionnelle (2) à la forme exempte du terme additif qui dépend de $x(T)$. Etant donné que

$$\frac{1}{2}\langle x(T),\ Fx(T)\rangle = \frac{1}{2}\langle x(t_0),\ Fx(t_0)\rangle + \int_{t_0}^{T}\frac{d}{dt}\left[\frac{1}{2}\langle x(t),\ Fx(t)\rangle\right]dt,$$

$$\frac{d}{dt}\left[\frac{1}{2}\langle x(t),\ Fx(t)\rangle\right] = \langle Fx(t),\ \dot{x}(t)\rangle,$$

on a

$$\frac{1}{2}\langle x(T),\ Fx(T)\rangle = \frac{1}{2}\langle x(t_0),\ Fx(t_0)\rangle + \int_{t_0}^{T}\langle Fx(t),\ \dot{x}(t)\rangle\,dt. \qquad (24.18)$$

En tenant compte de l'identité (18) on peut ramener la fonctionnelle (2) à la forme

$$\hat{J} = \frac{1}{2}\langle x(t_0),\ Fx(t_0)\rangle + \int_{t_0}^{T}\Big[\langle Fx(t),\ \dot{x}(t)\rangle + \frac{1}{2}\langle x(t),\ Q(t)\,x(t)\rangle + \frac{1}{2}\langle u(t),\ R(t)\,u(t)\rangle\Big]dt. \qquad (24.19)$$

La fonction $\mathscr{H}$ définie par l'expression (5) peut être remplacée respectivement par la fonction

$$\hat{\mathscr{H}} = \frac{1}{2}\langle x(t),\ Q(t)\,x(t)\rangle + \frac{1}{2}\langle u(t),\ R(t)\,u(t)\rangle + \langle Fx(t),\ \dot{x}(t)\rangle + \langle \hat{p},\ [A(t)\,x(t) + B(t)\,u(t)]\rangle.$$

Comme d'après (1)

$$\dot{x}(t) = A(t)\,x(t) + B(t)\,u(t),$$

la fonction $\hat{\mathscr{H}}$ s'écrit

$$\hat{\mathscr{H}} = \frac{1}{2}\langle x(t),\ Q(t)\,x(t)\rangle + \frac{1}{2}\langle u(t),\ R(t)\,u(t)\rangle + \langle [Fx(t) + \hat{p}],\ [A(t)\,x(t) + B(t)\,u(t)]\rangle. \qquad (24.20)$$

Puisque les fonctionnelles (2) et (19) se déduisent l'une de l'autre à l'aide de la transformation identique, les fonctions $\mathscr{H}$ et $\hat{\mathscr{H}}$ doivent coïncider. A cet effet, comme il résulte des expressions (5) et (20), il faut que la condition

$$p(t) = Fx(t) + \hat{p}(t) \qquad (24.21)$$

soit observée. Lorsque la fonctionnelle à minimiser est de la forme (19), les conditions de transversalité du problème à temps fixé et à extrémité libre de la trajectoire conduisent, d'après (17.55), à la relation

$$\hat{p}(T) = 0. \qquad (24.22)$$

Dans ces conditions, conformément à (21), le problème initial à fonctionnelle (2) donne lieu à la relation (17)

$$p(T) = Fx(T).$$

Cherchons encore l'équation différentielle vérifiée par le vecteur fonction $\hat{p}(t)$ du problème à fonctionnelle (19). D'une façon analogue à (6)

$$\frac{d}{dt}\hat{p}_j = -\frac{\partial\hat{\mathcal{H}}}{\partial x_j} \qquad (j=1, \ldots, n). \tag{24.23}$$

Comme

$$\frac{\partial}{\partial x_j}\langle Fx, Ax\rangle = \left\langle \begin{bmatrix} F_{1j} \\ F_{2j} \\ \ldots \\ F_{nj} \end{bmatrix}, Ax\right\rangle + \left\langle \begin{bmatrix} A_{1j} \\ A_{2j} \\ \ldots \\ A_{nj} \end{bmatrix}, Fx\right\rangle,$$

compte tenu du fait que F est une matrice symétrique, c'est-à-dire que $F^* = F$, on a

$$\operatorname{grad}\langle Fx, Ax\rangle = FAx + A^*Fx.$$

Le système d'équations différentielles (23) peut donc être mis sous la forme de l'équation différentielle vectorielle

$$\frac{d}{dt}\hat{p} = -Q(t)x - A^*(t)\hat{p} - F\dot{B}(t)u - FA(t)x - A^*(t)Fx \tag{24.24}$$

ou

$$\frac{d}{dt}\hat{p} = -Q(t)x - A^*(t)(Fx+\hat{p}) - F[A(t)x + B(t)u].$$

Puisque d'après (1)

$$\dot{x}(t) = A(t)x + B(t)u,$$

l'équation (24) se ramène à la forme

$$\frac{d}{dt}(\hat{p} + Fx) = -Q(t)x - A^*(t)(\hat{p} + Fx),$$

d'où, en vertu de (21), on obtient l'équation (7)

$$\dot{p} = -Q(t)x - A^*(t)p$$

du problème initial à fonctionnelle (2), ce qu'il fallait précisément démontrer.

Revenons maintenant aux équations différentielles (7) et (13)

$$\dot{p} = -Q(t)x - A^*(t)p,$$
$$\dot{x} = A(t)x - S(t)p.$$

Cherchons $p(t)$ sous la forme

$$p(t) = K(t)x(t), \tag{24.25}$$

où $K(t)$ est une matrice $n \times n$ à déterminer. Dans ce cas l'équation (13) devient

$$\dot{x} = [A(t) - S(t)K(t)]x. \tag{24.26}$$

La relation (25) entraîne

$$p(T) = K(T)x(T). \tag{24.27}$$

En comparant les expressions (27) et (17), on trouve que

$$K(T) = F. \tag{24.28}$$

Les équations (7) et (13) permettent d'obtenir l'équation différentielle vérifiée par la matrice $K(t)$. En portant dans ces équations l'expression (25), on obtient

$$\dot{K}(t)\,x(t) + K(t)\,A(t)\,\dot{x}(t) - K(t)\,S(t)\,K(t)\,x(t) = \\ = -Q(t)\,x(t) - A^*(t)\,K(t)\,x(t). \qquad (24.29)$$

Compte tenu du fait que d'après (14)

$$S(t) = B(t)\,R^{-1}(t)\,B^*(t),$$

l'équation (29) peut être ramenée à la forme

$$[\dot{K}(t) + K(t)\,A(t) + A^*(t)\,K(t) - \\ - K(t)\,B(t)\,R^{-1}(t)B^*(t)\,K(t) + Q(t)]\,x(t) = 0. \qquad (24.30)$$

La relation (30) doit être respectée quelle que soit la valeur de $x(t)$, d'où l'on déduit que la matrice $K(t)$ doit vérifier l'équation différentielle

$$\dot{K} = -KA(t) - A^*(t)\,K + KB(t)\,R^{-1}(t)\,B^*(t)\,K - Q(t) \\ (t \in [t_0,\ T]). \qquad (24.31)$$

L'expression (31) est une équation différentielle matricielle non linéaire de Riccati. Avec la condition aux limites (28), l'équation (31) détermine univoquement la matrice $K(t)$.

Montrons encore que la matrice $K(t)$ est symétrique. En transposant les matrices du premier et du second membre de (31) et en retenant que $B(t)\,R^{-1}(t)\,B^*(t)$ et $Q(t)$ sont des matrices symétriques et que

$$[\dot{K}(t)]^* = \frac{d}{dt}[K^*(t)],$$

on obtient l'équation

$$\frac{d}{dt}K^* = -K^*A(t) - A^*(t)\,K^* + \\ + K^*B(t)\,R^{-1}(t)\,B^*(t)\,K^* - Q(t) \quad (t_0 \in [t,\ T]). \qquad (24.32)$$

Comme la matrice F est symétrique, la condition aux limites (28) devient

$$K^*(T) = F. \qquad (24.33)$$

Ainsi les équations différentielles et les conditions aux limites vérifiées par $K(t)$ et $K^*(t)$ coïncident et, en vertu du théorème d'unicité,

$$K(t) = K^*(t), \qquad (24.34)$$

c'est-à-dire $K(t)$ est une matrice symétrique.

En vertu de (26) et (14), l'équation différentielle qui définit la loi du mouvement d'un système *optimal* est

$$\dot{x} = [A(t) - B(t) R^{-1}(t) B^*(t) K(t)]\, x. \tag{24.35}$$

Montrons maintenant comment la matrice $K(t)$ est liée à la matrice fondamentale $\hat{\theta}(t)$ des solutions de l'équation différentielle vectorielle (15). Désignons par $\Omega(t, \tau)$ la matrice $2n \times 2n$

$$\Omega(t, \tau) = \hat{\theta}(t)\, \hat{\theta}^{-1}(\tau). \tag{24.36}$$

L'équation (15) a pour solution

$$\begin{bmatrix} x(t) \\ p(t) \end{bmatrix} = \Omega(t, t_0) \begin{bmatrix} x(t_0) \\ p(t_0) \end{bmatrix}. \tag{24.37}$$

D'une façon analogue à (37), on a pour $T > t$

$$\begin{bmatrix} x(T) \\ p(T) \end{bmatrix} = \Omega(T, t) \begin{bmatrix} x(t) \\ p(t) \end{bmatrix}. \tag{24.38}$$

En mettant la matrice $\Omega(T, t)$ sous la forme d'une matrice partitionnée dont les éléments sont des matrices $n \times n$:

$$\Omega(T, t) = \begin{bmatrix} \Omega_{11}(T, t) & \Omega_{12}(T, t) \\ \Omega_{21}(T, t) & \Omega_{22}(T, t) \end{bmatrix}, \tag{24.39}$$

on obtient à partir de (38) et (17) les relations

$$\left.\begin{aligned} x(T) &= \Omega_{11}(T, t)\, x(t) + \Omega_{12}(T, t)\, p(t), \\ p(T) &= Fx(T) = \Omega_{21}(T, t)\, x(t) + \Omega_{22}(T, t)\, p(t). \end{aligned}\right\} \tag{24.40}$$

Des relations (40) résulte que

$$p(t) = \{[\Omega_{22}(T, t) - F\Omega_{12}(T, t)]^{-1} [F\Omega_{11}(T, t) - \Omega_{21}(T, t)]\}\, x(t). \tag{24.41}$$

Puisque $\Omega(T, T) = E_{2n}$, où E_{2n} est la matrice unité $2n \times 2n$, il vient

$$\Omega_{11}(T, T) = \Omega_{22}(T, T) = E_n,$$
$$\Omega_{12}(T, T) = \Omega_{21}(T, T) = 0,$$

où E_n est la matrice unité $n \times n$ et (41) conduit à

$$p(T) = Fx(T),$$

ce qui coïncide avec l'expression (17).

La comparaison des expressions (41) et (25) donne

$$K(t) = [\Omega_{22}(T, t) - F\Omega_{12}(T, t)]^{-1} [F\Omega_{11}(T, t) - \Omega_{21}(T, t)]. \tag{24.42}$$

La matrice $\Omega(T, t)$ peut être calculée par intégration de l'équation différentielle vectorielle conjuguée de l'équation différentielle vectorielle (15) [68].

Dans le cas où A, B, Q et R sont des matrices constantes, la construction de la matrice fondamentale $\hat{\theta}(t)$ ne présente aucune difficulté et la matrice $\Omega(T, t)$ peut être définie sous une forme fermée. Le système optimal décrit par l'équation différentielle vectorielle (35) est un système linéaire ; toutefois, même dans le cas où A, B, Q et R sont des matrices constantes, ce système optimal sera non stationnaire.

Passons maintenant à la détermination de la valeur minimale $J_1^*(x, t)$ de la fonctionnelle

$$J_1(x, t) = \frac{1}{2}\langle x(T), Fx(T)\rangle + \frac{1}{2}\int_t^T [\langle x(\xi), Q(\xi)x(\xi)\rangle + \\ + \langle u(\xi), R(\xi)u(\xi)\rangle]\, d\xi \quad (t \in [t_0, T]). \quad (24.43)$$

Montrons que

$$J_1^*(x, t) = \frac{1}{2}\langle x(t), K(t)x(t)\rangle \quad (t \in [t_0, T]), \quad (24.44)$$

où $K(t)$ est la solution de l'équation de Riccati (31) sous la condition aux limites (28). Dans ce cas, compte tenu de (28), on a

$$J_1^*(x(T), T) = \frac{1}{2}\langle x(T), Fx(T)\rangle. \quad (24.45)$$

Pour montrer la validité de l'expression (44), composons l'équation de Bellman pour le problème considéré. D'après (14.21), cette équation s'écrit

$$\frac{\partial J_1^*}{\partial t} + \min_{u(t)}\Big\{\frac{1}{2}\langle x(t), Q(t)x(t)\rangle + \frac{1}{2}\langle u(t), R(t)u(t)\rangle + \\ + \langle A(t)x(t), \operatorname{grad} J_1^*(x, t)\rangle + \langle B(t)u(t), \operatorname{grad} J_1^*(x, t)\rangle\Big\} = 0. \quad (24.46)$$

L'expression entre accolades de l'équation (46) coïncide avec l'expression (5) qui définit la fonction $\mathscr{H}$ si l'on y remplace le vecteur $p(t)$ par le vecteur $\operatorname{grad} J_1^*(x, t)$. Conformément à (11), la commande minimisant cette expression est

$$u(t) = -R^{-1}(t)B^*(t)\operatorname{grad} J_1^*(x, t). \quad (24.47)$$

Comme d'après (44)

$$\frac{\partial J_1^*}{\partial t} = \frac{\partial}{\partial t}\left[\frac{1}{2}\langle x, K(t)x\rangle\right] = \frac{1}{2}\langle x, \dot K(t)x\rangle,$$

$$\operatorname{grad} J_1^* = \operatorname{grad}\left[\frac{1}{2}\langle x, K(t)x\rangle\right] = K(t)x,$$

la formule (47) se met sous la forme

$$u(t) = -R^{-1}(t)B^*(t)K(t)x(t). \quad (24.48)$$

L'équation (46) peut s'écrire

$$\frac{1}{2}\langle x(t), \dot K(t)x(t)\rangle + \frac{1}{2}\langle x(t), Q(t)x(t)\rangle + \\ + \frac{1}{2}\langle R^{-1}(t)B^*(t)K(t)x(t), B^*(t)K(t)x(t)\rangle + \\ + \langle A(t)x(t), K(t)x(t)\rangle - \\ - \langle B(t)R^{-1}(t)B^*(t)K(t)x(t), K(t)x(t)\rangle = 0. \quad (24.49)$$

Pour transformer les équations (49) on recourt aux identités

$$\langle R^{-1}B^*Kx,\ B^*Kx\rangle = (R^{-1}B^*Kx)^*\ B^*Kx = \\ = (BR^{-1}B^*Kx)^*\ Kx = \langle BR^{-1}B^*Kx,\ Kx\rangle, \qquad (24.50)$$

$$\langle x,\ KAx\rangle = x^*KAx = (K^*x)^*\ Ax = \langle K^*x,\ Ax\rangle = \langle Kx,\ Ax\rangle = \langle Ax,\ Kx\rangle.$$

Nous avons tenu compte ici du fait que d'après (34) $K = K^*$. D'une façon analogue

$$\langle x,\ A^*Kx\rangle = x^*A^*Kx = (Ax)^*\ Kx = \langle Ax,\ Kx\rangle.$$

Il suit des deux dernières identités que

$$\langle Ax,\ Kx\rangle = \frac{1}{2}\langle x,\ A^*Kx\rangle + \frac{1}{2}\langle x,\ KAx\rangle. \qquad (24.51)$$

A l'aide de (50) et (51) on peut réduire l'équation (49) à la forme

$$\frac{1}{2}\langle x(t),\ \dot{K}(t)\,x(t)\rangle + \frac{1}{2}\langle x(t),\ Q(t)\,x(t)\rangle + \\ + \frac{1}{2}\langle x(t),\ A^*(t)\,K(t)\,x(t)\rangle + \frac{1}{2}\langle x(t),\ K(t)\,A(t)\,x(t)\rangle - \\ - \frac{1}{2}\langle x(t),\ K(t)\,B(t)\,R^{-1}(t)\,B^*(t)\,K(t)\,x(t)\rangle = 0. \qquad (24.52)$$

Ainsi, conformément à (44), l'équation de Bellman (46) devient

$$\frac{1}{2}\langle x(t),\ [\dot{K}(t) + K(t)\,A(t) + A^*(t)\,K(t) - \\ - K(t)\,B(t)\,R^{-1}(t)\,B^*(t)\,K(t) + Q(t)]\,x(t)\rangle = 0. \qquad (24.53)$$

Comme $K(t)$ est une solution de l'équation de Riccati (31), l'expression entre crochets de l'équation (53) est identiquement nulle et, par suite, la fonction $J_1^*(x,\ t)$ déterminée par l'expression (44) est, dans notre problème, une solution de l'équation de Bellman.

Ce qui vient d'être dit montre que la fonction $J_1^*(x,\ t)$ vérifie l'équation de Bellman (53) quelles que soient les valeurs de ses variables: $x(t) \in X$ (où X est un espace de phase de dimension n) et $t \in [t_0,\ T]$. On peut prouver [14] (d'une façon analogue au théorème du § 15) qu'il s'ensuit que la commande (48) satisfait aux conditions nécessaires et suffisantes d'optimalité et que cette commande est unique.

Montrons encore que pour $t \in [t_0,\ T)$ la matrice $K(t)$ est définie positive. F et $Q(t)$ étant des matrices définies non négatives, et $R(t)$ une matrice définie positive, d'après (43) la fonctionnelle $J_1(x,\ t)$ présente, quelle que soit la commande $u(t) \not\equiv 0$, une valeur positive pour tout $t \in [t_0,\ T)$. Par conséquent, pour tout vecteur $x(t)$ la valeur minimale $J_1^*(x,\ t)$ de la fonctionnelle $J_1(x,\ t)$ définie par l'expression (44) est

$$J_1^*(x,\ t) = \frac{1}{2}\langle x(t),\ K(t)\,x(t)\rangle > 0, \qquad t \in [t_0,\ T).$$

Il en résulte que la matrice $K(t)$ est définie positive pour $t \in [t_0,\ T)$.

Calcul de la valeur limite de $p(T)$ pour le cas général du problème à temps fixé et à extrémité libre de la trajectoire. Dans le cas général l'équation du mouvement du système s'écrit

$$\dot{x} = f(x, u, t), \tag{24.54}$$

où f et x sont des vecteurs de dimension n, et u le vecteur de dimension r.

Soit

$$J = N(x(T), T) + \int_{t_0}^{T} L(x(t), u(t), t)\, dt \tag{24.55}$$

la fonctionnelle à minimiser. D'une façon analogue à (5), la fonction $\mathcal{H}$ à minimiser par la commande optimale est de la forme

$$\mathcal{H} = L(x, u, t) + \langle p, f(x, u, t)\rangle. \tag{24.56}$$

Comme

$$N(x(T), T) = N(x(t_0), t_0) + \int_{t_0}^{T} \frac{d}{dt} N(x(t), t)\, dt,$$

$$\frac{d}{dt} N(x(t), t) = \langle \operatorname{grad} N, \dot{x}\rangle + \frac{\partial N}{\partial t} = \langle \operatorname{grad} N, f(x, u, t)\rangle + \frac{\partial N}{\partial t},$$

on a

$$N(x(T), T) = N(x(t_0), t_0) + \int_{t_0}^{T} \left[\langle \operatorname{grad} N, f(x, u, t)\rangle + \frac{\partial N}{\partial t}\right] dt. \tag{24.57}$$

La fonctionnelle (55) se réduit à la forme

$$\hat{J} = N(x(t_0), t_0) + \int_{t_0}^{T} \left[\langle \operatorname{grad} N, f(x, u, t)\rangle + \frac{\partial N}{\partial t} + L(x, u, t)\right] dt. \tag{24.58}$$

Maintenant il faut remplacer la fonction $\mathcal{H}$ par la fonction

$$\hat{\mathcal{H}} = \langle \operatorname{grad} N, f(x, u, t)\rangle + \frac{\partial N}{\partial t} + L(x, u, t) + \langle \hat{p}, f(x, u, t)\rangle. \tag{24.59}$$

Cherchons l'équation différentielle qui doit être vérifiée par le vecteur fonction $\hat{p}(t)$. Puisque

$$\frac{d\hat{p}_j}{dt} = -\frac{\partial \hat{\mathcal{H}}}{\partial x_j} \quad (j = 1, \ldots, n)$$

on a 168

$$\frac{d\hat{p}_j}{dt} = -\frac{\partial}{\partial x_j} \langle \operatorname{grad} N, f(x, u, t)\rangle - \frac{\partial^2 N}{\partial x_j\, \partial t} - \frac{\partial L}{\partial x_j} - \frac{\partial}{\partial x_j} \langle \hat{p}, f(x, u, t)\rangle$$

ou

$$\frac{d\hat{p}_j}{dt} = -\left\langle \frac{\partial}{\partial x_j} \operatorname{grad} N, f(x, u, t)\right\rangle - \left\langle \operatorname{grad} N, \frac{\partial f(x, u, t)}{\partial x_j}\right\rangle - \\ - \frac{\partial^2 N}{\partial x_j\, \partial t} - \frac{\partial L}{\partial x_j} - \left\langle \hat{p}, \frac{\partial f(x, u, t)}{\partial x_j}\right\rangle. \tag{24.60}$$

Montrons que le vecteur $p(t)$ qui intervient dans l'expression (56) est

$$p(t)=\hat{p}(t)+\operatorname{grad} N(x,\ t), \tag{24.61}$$

c'est-à-dire que

$$p_j(t)=\hat{p}_j(t)+\frac{\partial N(x,\ t)}{\partial x_j} \qquad (j=1,\ \ldots,\ n). \tag{24.62}$$

On tire de (62) et (60)

$$\dot{p}_j=-\left\langle\frac{\partial}{\partial x_j}\operatorname{grad} N,\ f(x,\ u,\ t)\right\rangle-\left\langle\operatorname{grad} N,\ \frac{\partial f(x,\ u,\ t)}{\partial x_j}\right\rangle-$$
$$-\frac{\partial^2 N}{\partial x_j\,\partial t}-\frac{\partial L}{\partial x_j}-\left\langle(p-\operatorname{grad} N),\ \frac{\partial f(x,\ u,\ t)}{\partial x_j}\right\rangle+\frac{d}{dt}\frac{\partial N}{\partial x_j}. \tag{24.63}$$

Puisque

$$\frac{d}{dt}\frac{\partial N}{\partial x_j}=\left\langle\frac{\partial}{\partial x_j}\operatorname{grad} N,\ \dot{x}\right\rangle+\frac{\partial^2 N}{\partial x_j\,\partial t}=$$
$$=\left\langle\frac{\partial}{\partial x_j}\operatorname{grad} N,\ f(x,\ u,\ t)\right\rangle+\frac{\partial^2 N}{\partial x_j\,\partial t}, \tag{24.64}$$

l'équation (63) devient

$$\dot{p}_j=-\frac{\partial L}{\partial x_j}-\left\langle p,\ \frac{\partial f(x,\ u,\ t)}{\partial x_j}\right\rangle \qquad (j=1,\ \ldots,\ n), \tag{24.65}$$

soit

$$\dot{p}_j=-\frac{\partial\mathscr{H}}{\partial x_j} \qquad (j=1,\ \ldots,\ n), \tag{24.66}$$

où la fonction $\mathscr{H}$ est déterminée par l'expression (56). Ce résultat montre que la relation (61) est vraie.

Lorsque dans un problème à temps fixé et à extrémité libre de la trajectoire la fonctionnelle à minimiser est de la forme (58), d'après (17.55) les conditions de transversalité conduisent à la relation

$$\hat{p}(T)=0. \tag{24.67}$$

De plus, conformément à (61), pour le problème initial à fonctionnelle (55) on a la relation

$$p(T)=\operatorname{grad} N(x(T),\ T). \tag{24.68}$$

2. Problème du régulateur de sortie. Considérons le système linéaire non stationnaire

$$\left.\begin{aligned}\dot{x}&=A(t)\,x+B(t)\,u(t),\\ y&=C(t)\,x,\end{aligned}\right\} \tag{24.69}$$

où x est le vecteur de dimension n des coordonnées de phase du système (c'est-à-dire son vecteur d'état) ; $u(t)$, le vecteur de dimension r de la commande ; $y(t)$, le vecteur de dimension m qui détermine la sortie du système ; $A(t)$, la matrice $n\times n$; $B(t)$, la matrice $n\times r$; $C(t)$, la matrice $m\times n$. On suppose que le système (69) est complète-

ment observable, c'est-à-dire que la matrice

$$M(t_0, T) = \int_{t_0}^{T} \Phi^*(t, t_0) C^*(t) C(t) \Phi(t, t_0)\, dt \qquad (24.70)$$

est définie positive. Ici

$$\Phi(t, \tau) = \theta(t)\, \theta^{-1}(\tau),$$

où $\theta(t)$ est la matrice fondamentale des solutions de l'équation différentielle vectorielle

$$\dot{x} = A(t)\, x.$$

De même qu'au point 1, on suppose que les commandes $u_j(t)$ ($j = 1, \ldots, r$) ne subissent pas de contraintes. On demande de choisir le vecteur de commande $u(t)$ tel qu'il minimise la fonctionnelle

$$J_2 = \frac{1}{2}\langle y(T), Fy(T)\rangle + \frac{1}{2}\int_{t_0}^{T} [\langle y(t), Q(t)\, y(t)\rangle + \\ + \langle u(t), R(t)\, u(t)\rangle]\, dt, \qquad (24.71)$$

où T est un instant fixé. Le but de la commande est de maintenir le vecteur $y(t)$ dans le voisinage du zéro, c'est-à-dire le problème consiste à régler la *sortie* du système. Les matrices F et $Q(t)$ sont des matrices définies non négatives $m \times m$, et la matrice $R(t)$ est une matrice définie positive $r \times r$. Comme

$$\langle y(t), Q(t)\, y(t)\rangle = \langle C(t)\, x(t), Q(t)\, C(t)\, x(t)\rangle = \\ = \langle x(t), C^*(t)\, Q(t)\, C(t)\, x(t)\rangle, \\ \langle y(T), Fy(T)\rangle = \langle x(T), C^*(T)\, FC(T)\, x(T)\rangle,$$

la fonctionnelle (71) peut s'écrire

$$J_2 = \frac{1}{2}\langle x(T), C^*(T)\, FC(T)\, x(T)\rangle + \\ + \frac{1}{2}\int_{t_0}^{T} \langle x(t), C^*(t)\, Q(t)\, C(t)\, x(t)\rangle + \langle u(t), R(t)\, u(t)\rangle]\, dt. \qquad (24.72)$$

Constatons que puisque F et $Q(t)$ sont des matrices définies non négatives, les matrices symétriques $C^*(T)\, FC(T)$ et $C^*(t)\, Q(t)\, C(t)$ sont également des matrices définies non négatives.

Par suite,

$$\langle x(T), C^*(T)\, FC(T)\, x(T)\rangle \geqslant 0 \qquad (24.73)$$

$$\langle x(t), C^*(t)\, Q(t)\, C(t)\, x(t)\rangle \geqslant 0. \qquad (24.74)$$

Si le système est complètement observable, on peut déterminer d'une façon unique d'après le vecteur fonction $y(t)$ dans $[0, T]$ le vecteur initial $x(t_0)$ du système (69), alors que tout vecteur $x(t_0)$ définit une trajectoire unique $x(t)$ de (69). Par conséquent, les équations (69) d'un système complètement observable définissent une application biunivoque de l'espace Y dans l'espace X et la condition d'observabilité complète assure ainsi l'unicité de la commande optimale d'un problème de régulation de la sortie d'un système.

La différence entre les fonctionnelles (2) et (72) ne portant que sur le fait que les matrices définies non négatives F et $Q(t)$ sont remplacées par les matrices définies non négatives $C^*(T)FC(T)$ et $C^*(t)Q(t)C(t)$ respectivement, les résultats du point 1 peuvent être appliqués au problème examiné.

D'une façon analogue à (48), la commande optimale est de la forme

$$u(t) = -R^{-1}(t)B^*(t)\mathscr{K}(t)x(t). \tag{24.75}$$

L'équation différentielle régissant le mouvement d'un système optimal est, d'après (35) et (69),

$$\dot{x} = [A(t) - B(t)R^{-1}(t)B^*(t)\mathscr{K}(t)]x. \tag{24.76}$$

La valeur minimale $J_2^*(x, t)$ de la fonctionnelle $J_2(x, t)$, définie d'une façon analogue à (43), est

$$J_2^*(x, t) = \frac{1}{2}\langle x(t), \mathscr{K}(t)x(t)\rangle, \quad t \in [t_0, T]. \tag{24.77}$$

La matrice symétrique définie positive $\mathscr{K}(t)$ qui intervient dans les expressions (75) et (77) et dans l'équation différentielle (76) est la solution de l'équation de Riccati

$$\dot{\mathscr{K}} = -\mathscr{K}A(t) - A^*(t)\mathscr{K} + \mathscr{K}B(t)R^{-1}(t)B^*(t)\mathscr{K} - \\ -C^*(t)Q(t)C(t) \qquad (t \in [t_0, T]), \tag{24.78}$$

qui vérifie la condition aux limites

$$\mathscr{K}(T) = C^*(T)FC(T). \tag{24.79}$$

3. Systèmes stationnaires à temps infini d'observation. a) *Régulation de l'état d'un système.* L'équation du mouvement du système s'écrit

$$\dot{x} = Ax + Bu.$$

Considérons le cas limite où l'extrémité de l'intervalle d'observation $T \to \infty$. Dans ces conditions, la fonctionnelle à minimiser s'écrit

$$J_\infty = \frac{1}{2}\lim_{T\to\infty}\int_0^T \langle x(t), Qx(t)\rangle + \langle u(t), Ru(t)\rangle]\,dt,$$

où Q et R sont des matrices définies positives constantes. La commande $u(t)$ n'est soumise à aucune contrainte. On suppose que le système est *complètement commandable*, soit

$$\text{rang}\,[B\ AB\ \ldots\ A^{n-1}B] = n.$$

L'équation de Riccati (31) devient

$$\dot{K} = -KA - A^*K + KBR^{-1}B^*K - Q.$$

Étant donné que dans la fonctionnelle J_∞, contrairement à la fonctionnelle (2), $F = 0$, la condition aux limites (28) peut s'écrire

$$K(T) = 0.$$

La solution de l'équation de Riccati vérifiant la condition aux limites $K(T) = 0$ se note $K(t; 0, T)$. Comme il suit des résultats fournis par Kalman [34], dans le cas des systèmes stationnaires complètement *commandables*, pour $F = 0$, cette solution possède une limite pour $T \to \infty$

$$\lim_{T\to\infty} K(t; 0, T) = \hat{K},$$

où $\hat{K}$, matrice définie positive symétrique constante, est la solution de l'équation algébrique matricielle non linéaire

$$-\hat{K}A - A^*\hat{K} + \hat{K}BR^{-1}B^*\hat{K} - Q = 0.$$

D'une façon analogue à (48), la commande optimale est de la forme

$$u(t) = -R^{-1}B^*\hat{K}x(t).$$

L'équation différentielle du mouvement du système optimal est

$$\dot{x} = (A - BR^{-1}B^*\hat{K})\,x,$$

de plus, le système optimal est *asymptotiquement stable*, c'est-à-dire

$$\lim_{t\to\infty} x(t) = 0.$$

La valeur minimale $J_\infty^*(x, t)$ de la fonctionnelle

$$J_\infty(x, t) = \frac{1}{2}\int_t^\infty [\langle x(\xi), Qx(\xi)\rangle + \langle u(\xi), Ru(\xi)\rangle]\,d\xi$$

est, de façon analogue à (44),

$$J_\infty(x; t) = \frac{1}{2}\langle x(t), \hat{K}x(t)\rangle.$$

b) Régulation de la sortie d'un système. Un système est décrit par les équations

$$\dot{x} = Ax + Bu,$$
$$y = Cx.$$

On demande de minimiser la fonctionnelle

$$\mathcal{J}_\infty = \frac{1}{2} \lim_{T\to\infty} \int_0^T [\langle y(t), Qy(t)\rangle + \langle u(t), Ru(t)\rangle]\, dt,$$

où Q et R sont des matrices définies positives constantes. La commande $u(t)$ ne subit aucune contrainte. On suppose que le système est comlètement commandable et complètement observable, soit

$$\operatorname{rang}[B\; AB \;\ldots\; A^{n-1}B] = n, \quad \operatorname{rang}[C^*\; A^*C^* \;\ldots\; A^{*n-1}C^*] = n.$$

D'une façon analogue au problème précédent, au cas limite pour $T \to \infty$ la commande optimale est

$$u(t) = -R^{-1}B^*\hat{\mathscr{K}}x(t),$$

où la matrice définie positive symétrique constante $\hat{\mathscr{K}}$ est la solution de l'équation algébrique matricielle non linéaire

$$-\hat{\mathscr{K}}A - A^*\hat{\mathscr{K}} + \hat{\mathscr{K}}BR^{-1}B^*\hat{\mathscr{K}} - C^*QC = 0.$$

L'équation du mouvement du système optimal

$$\dot{x} = (A - BR^{-1}B^*\hat{\mathscr{K}})\, x.$$

Le système optimal est *asymptotiquement stable*, c'est-à-dire

$$\lim_{t\to\infty} x(t) = 0.$$

La valeur minimale $\tilde{\mathcal{J}}^*_\infty(x, t)$ de la fonctionnelle

$$\tilde{\mathcal{J}}_\infty(x, t) = \frac{1}{2} \int_t^\infty [\langle Cx(\xi), QCx(\xi)\rangle + \langle u(\xi), Ru(\xi)\rangle]\, d\xi$$

est

$$\tilde{\mathcal{J}}^*_\infty(x, t) = \frac{1}{2} \langle x(t), \hat{\mathscr{K}}x(t)\rangle.$$

4. **Problème d'asservissement [34].** Examinons le système linéaire non stationnaire

$$\left.\begin{aligned} \dot{x} &= A(t)\,x + B(t)\,u(t), \\ y &= C(t)\,x, \end{aligned}\right\} \qquad (24.80)$$

où x est le vecteur d'état de dimension n ; $u(t)$, le vecteur de commande de dimension r; $y(t)$, le vecteur de dimension m qui détermine la sortie du système; $A(t)$, la matrice $n \times n$; $B(t)$, la matrice $n \times r$; $C(t)$, la matrice $m \times n$. On suppose que le système (80) est complètement observable, c'est-à-dire que la matrice $M(t_0, T)$ du type $n \times n$ déterminée par l'expression (70) est une matrice définie positive.

Désignons par $z(t)$ le vecteur de dimension m muni d'une loi de variation dans le temps donnée de l'extérieur et qui présente la sortie *cherchée* du système. Le vecteur

$$e(t) = z(t) - y_1(t) \tag{24.81}$$

est l'écart ou l'erreur du système asservi.

Sous l'hypothèse que la commande $u(t)$ n'est soumise à aucune contrainte, on demande de minimiser la fonctionnelle

$$J_3 = \frac{1}{2}\langle e(T), Fe(T)\rangle + \\ + \frac{1}{2}\int_{t_0}^{T} [\langle e(t), Q(t)e(t)\rangle + \langle u(t), R(t)u(t)\rangle]\,dt, \tag{24.82}$$

où T est un instant fixé. Ainsi, le but de la commande est de maintenir l'écart du système, c'est-à-dire le vecteur $e(t)$, dans le voisinage du zéro. Les matrices F et $Q(t)$ sont des matrices définies non négatives $m \times m$, et la matrice $R(t)$ une matrice définie positive $r \times r$.

En remplaçant dans (82) $e(t)$ par l'expression

$$e(t) = z(t) - C(t)x(t), \tag{24.83}$$

ramenons la fonctionnelle (82) à la forme

$$J_3 = \frac{1}{2}\langle [z(T) - C(T)x(T)], F[z(T) - C(T)x(T)]\rangle + \\ + \frac{1}{2}\int_{t_0}^{T} \{\langle [z(t) - C(t)x(t)], Q(t)[z(t) - C(t)x(t)]\rangle + \\ + \langle u(t), R(t)u(t)\rangle\}\,dt. \tag{24.84}$$

La commande optimale $u(t)$ doit minimiser la fonction $\tilde{\mathscr{H}}$ qui se construit d'une façon analogue à la fonction $\mathscr{H}$ définie par l'expression (5). La fonction $\tilde{\mathscr{H}}$ associée à la fonctionnelle (84) s'écrit

$$\tilde{\mathscr{H}} = \frac{1}{2}\langle [z(t) - C(t)x(t)], Q(t)[z(t) - C(t)x(t)]\rangle + \\ + \frac{1}{2}\langle u(t), R(t)u(t)\rangle + \langle A(t)x(t), p(t)\rangle + \\ + \langle B(t)u(t), p(t)\rangle. \tag{24.85}$$

Les relations (8)-(11) conservent leur forme pour le problème envisagé également et conformément à (11), la commande $u(t)$ minimisant la fonction $\tilde{\mathscr{H}}$ est

$$u(t) = -R^{-1}(t)B^*(t)p(t). \tag{24.86}$$

En portant l'expression (86) de $u(t)$ dans la première équation (80), on met cette équation sous la forme

$$\dot{x} = A(t)\,x - S(t)\,p, \tag{24.87}$$

où

$$S(t) = B(t)\,R^{-1}(t)\,B^*(t). \tag{24.88}$$

Comme

$$\dot{p}_j = -\frac{\partial \mathscr{H}}{\partial x_j} \qquad (j = 1, \ldots, n),$$

le vecteur fonction $p(t)$ vérifie l'équation différentielle

$$\dot{p} = -C^*(t)\,Q(t)\,C(t)\,x - A^*(t)\,p + C^*(t)\,Q(t)\,z(t). \tag{24.89}$$

En introduisant les notations

$$V(t) = C^*(t)\,Q(t)\,C(t), \quad W(t) = C^*(t)\,Q(t), \tag{24.90}$$

on peut écrire (89) sous la forme

$$\dot{p} = -V(t)\,x - A^*(t)\,p + W(t)\,z(t). \tag{24.91}$$

Les équations (87) et (91) composent un système d'équations différentielles vectorielles qui peut se noter

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} A(t) & -S(t) \\ -V(t) & -A^*(t) \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix} + \begin{bmatrix} 0 \\ W(t)\,z(t) \end{bmatrix}. \tag{24.92}$$

La solution du système d'équations différentielles (92) de rang $2n$ doit vérifier n conditions initiales définies par la relation vectorielle

$$[x(t)]_{t=t_0} = x(t_0). \tag{24.93}$$

Conformément à (68), (55) et (82), les conditions aux limites des variables auxiliaires p_j $(j = 1, \ldots, n)$ sont déterminées par la relation

$$p_j(T) = \left\{ \frac{\partial}{\partial x_j} \left[\frac{1}{2} \langle e(t), Fe(t) \rangle \right] \right\}_{t=T}. \tag{24.94}$$

Puisque

$$\frac{1}{2}\langle e(t), Fe(t)\rangle = \frac{1}{2}\langle z(t), Fz(t)\rangle -$$
$$- \langle z(t), FC(t)\,x(t)\rangle + \frac{1}{2}\langle x(t), C^*(t)\,FC(t)\,x(t)\rangle,$$

d'après (94), pour le vecteur fonction $p(t)$ on a la condition aux limites

$$p(T) = C^*(T)\,FC(T)\,x(T) - C^*(T)\,Fz(T). \tag{24.95}$$

Désignons maintenant par $\widetilde{\theta}(t)$ la matrice fondamentale des solutions de l'équation différentielle vectorielle homogène qui se construit à partir de l'équation (92) avec $z(t) \equiv 0$. Notons $\widetilde{\Omega}(t, \tau)$ la matrice $2n \times 2n$

$$\widetilde{\Omega}(t, \tau) = \widetilde{\theta}(t)\,\widetilde{\theta}^{-1}(\tau). \tag{24.96}$$

La solution de l'équation (92) s'écrit

$$\begin{bmatrix} x(t) \\ p(t) \end{bmatrix} = \widetilde{\Omega}(t, t_0) \begin{bmatrix} x(t_0) \\ p(t_0) \end{bmatrix} + \int_{t_0}^{t} \widetilde{\Omega}(t, \tau) \begin{bmatrix} 0 \\ W(\tau) z(\tau) \end{bmatrix} d\tau. \quad (24.97)$$

D'une façon analogue à (97) on a pour $T > t$

$$\begin{bmatrix} x(T) \\ p(T) \end{bmatrix} = \widetilde{\Omega}(T, t) \begin{bmatrix} x(t) \\ p(t) \end{bmatrix} + \int_{t}^{T} \widetilde{\Omega}(T, \tau) \begin{bmatrix} 0 \\ W(\tau) z(\tau) \end{bmatrix} d\tau \qquad (t \in [t_0, T]). \quad (24.98)$$

En représentant la matrice $\widetilde{\Omega}(T, t)$ du type $2n \times 2n$ sous une forme partitionnée dont les éléments sont des matrices $n \times n$

$$\widetilde{\Omega}(T, t) = \begin{bmatrix} \widetilde{\Omega}_{11}(T, t) & \widetilde{\Omega}_{12}(T, t) \\ \widetilde{\Omega}_{21}(T, t) & \widetilde{\Omega}_{22}(T, t) \end{bmatrix}, \quad (24.99)$$

on peut réduire (98) à la forme

$$\begin{bmatrix} x(T) \\ p(T) \end{bmatrix} = \begin{bmatrix} \widetilde{\Omega}_{11}(T, t) x(t) + \widetilde{\Omega}_{12}(T, t) p(t) \\ \widetilde{\Omega}_{21}(T, t) x(t) + \widetilde{\Omega}_{22}(T, t) p(t) \end{bmatrix} + \begin{bmatrix} \Gamma_1(t) \\ \Gamma_2(t) \end{bmatrix}, \quad (24.100)$$

où

$$\Gamma_j(t) = \int_{t}^{T} \widetilde{\Omega}_{j2}(T, \tau) W(\tau) z(\tau) d\tau \qquad (j = 1, 2). \quad (24.101)$$

Il s'ensuit de (100) que

$$x(T) = \widetilde{\Omega}_{11}(T, t) x(t) + \widetilde{\Omega}_{12}(T, t) p(t) + \Gamma_1(t), \quad (24.102)$$

$$p(T) = \widetilde{\Omega}_{21}(T, t) x(t) + \widetilde{\Omega}_{22}(T, t) p(t) + \Gamma_2(t). \quad (24.103)$$

En substituant dans (103) à $p(T)$ l'expression (95), on obtient la relation

$$\begin{aligned} [C^*(T) FC(T) \widetilde{\Omega}_{12}(T, t) - \widetilde{\Omega}_{22}(T, t)] p(t) = \\ = [-C^*(T) FC(T) \widetilde{\Omega}_{11}(T, t) + \widetilde{\Omega}_{21}(T, t)] x(t) - \\ - C^*(T) FC(T) \Gamma_1(t) + \Gamma_2(t) + C^*(T) Fz(T). \end{aligned} \quad (24.104)$$

Puisque

$$\widetilde{\Omega}_{11}(T, T) = \widetilde{\Omega}_{22}(T, T) = E,$$
$$\widetilde{\Omega}_{12}(T, T) = \widetilde{\Omega}_{21}(T, T) = 0,$$
$$\Gamma_1(T) = 0, \ \Gamma_2(T) = 0,$$

pour $t = T$ la relation (104) se met sous la forme (95)

$$p(T) = C^*(T) FC(T) x(T) - C^*(T) Fz(T).$$

La relation (104) entraîne que le vecteur fonction $p(t)$ peut s'écrire

$$p(t) = K(t) x(t) - g(t), \quad (24.105)$$

de plus, en vertu de (95),

$$K(T) = C^*(T)\,FC(T), \tag{24.106}$$

$$g(T) = C^*(T)\,Fz(T). \tag{24.107}$$

Cherchons maintenant les équations différentielles vérifiées par la matrice $K(t)$ et le vecteur $g(t)$. D'après (105), on a

$$\dot{p}(t) = \dot{K}(t)\,x(t) + K(t)\,\dot{x}(t) - \dot{g}(t). \tag{24.108}$$

Conformément à (87) et à (105), l'équation (108) peut être ramenée à la forme

$$\dot{p}(t) = \dot{K}(t)\,x(t) + K(t)\,A(t)\,x(t) - \\ - K(t)\,S(t)\,K(t)\,x(t) + K(t)\,S(t)\,g(t) - \dot{g}(t). \tag{24.109}$$

En éliminant $\dot{p}$ des équations (109) et (91) et en remplaçant $p(t)$ par l'expression (105), on obtient la relation

$$[\dot{K}(t) + K(t)\,A(t) + A^*(t)\,K(t) - K(t)\,S(t)\,K(t) + \\ + V(t)]\,x(t) - \{\dot{g}(t) - [K(t)\,S(t) - A^*(t)]\,g(t) + \\ + W(t)\,z(t)\} = 0. \tag{24.110}$$

La relation (110) ayant lieu quels que soient $x(t)$, $z(t)$ et t, il faut que soient remplies les conditions suivantes:

$$\dot{K}(t) + K(t)\,A(t) + A^*(t)\,K(t) - K(t)\,S(t)\,K(t) + V(t) = 0, \tag{24.111}$$

$$\dot{g}(t) - [K(t)\,S(t) - A^*(t)]\,g(t) + W(t)\,z(t) = 0, \tag{24.112}$$

c'est-à-dire $K(t)$ et $g(t)$ doivent être des solutions des équations différentielles (111) et (112).

En remplaçant $S(t)$, $V(t)$ et $W(t)$ par les expressions (88) et (90), on obtient que la matrice $K(t)$ est une solution de l'équation différentielle de Riccati

$$\dot{K} = -KA(t) - A^*(t)\,K + KB(t)\,R^{-1}(t)\,B^*(t)\,K - \\ - C^*(t)\,Q(t)\,C(t) \qquad (t \in [t_0, T]), \tag{24.113}$$

qui vérifie la condition aux limites (106)

$$K(T) = C^*(T)\,FC(T).$$

Le vecteur fonction $g(t)$ est la solution de l'équation différentielle

$$\dot{g} = -[A(t) - B(t)\,R^{-1}(t)\,B^*(t)\,K(t)]^*\,g - \\ - C^*(t)\,Q(t)\,z(t) \qquad (t \in [t_0, T]), \tag{24.114}$$

qui satisfait à la condition aux limites (107)

$$g(T) = C^*(T)\,Fz(T).$$

L'équation différentielle qui, conformément à (87), (88) et (105) décrit la loi du mouvement d'un système optimal, est de la forme

$$\dot{x} = [A(t) - B(t) R^{-1}(t) B^*(t) K(t)]\, x + \\ + B(t) R^{-1}(t) B^*(t) g(t) \qquad (t \in [t_0, T]). \qquad (24.115)$$

D'après (86) et (105), la commande optimale $u(t)$ s'écrit

$$u(t) = -R^{-1}(t) B^*(t) K(t) x(t) + R^{-1}(t) B^*(t) g(t). \qquad (24.116)$$

Notons que l'équation différentielle (113) ne possède pas de termes dépendant du signal d'entrée $z(t)$. Ainsi, la matrice $K(t)$ ne dépend pas du signal d'entrée $z(t)$.

Pour déterminer d'après (116) la commande optimale $u(t)$ ainsi que pour intégrer l'équation différentielle (115), il faut connaître la fonction $g(t)$, ce qui exige l'intégration de l'équation différentielle (114). Désignons par $\vartheta(t)$ la matrice fondamentale des solutions de l'équation homogène qui se construit à partir de l'équation (114) avec $z(t) \equiv 0$. Introduisons la notation $\psi(t, \tau)$ pour la matrice

$$\psi(t, \tau) = \vartheta(t)\, \vartheta^{-1}(\tau). \qquad (24.117)$$

La solution de l'équation vectorielle (114) s'écrit alors

$$g(T) = \psi(T, t)\, g(t) - \int_t^T \psi(T, \tau)\, C^*(\tau)\, Q(\tau)\, z(\tau)\, d\tau \qquad (24.118)$$

$$(t \in [t_0, T]).$$

Comme

$$\psi^{-1}(T, t) = \psi(t, T), \quad \psi(t, T)\, \psi(T, \tau) = \psi(t, \tau),$$

(118) conduit à

$$g(t) = \psi(t, T)\, g(T) + \int_t^T \psi(t, \tau)\, C^*(\tau)\, Q(\tau)\, z(\tau)\, d\tau,$$

ou, en vertu de (107),

$$g(t) = \psi(t, T)\, C^*(T)\, F z(T) + \int_t^T \psi(t, \tau)\, C^*(\tau)\, Q(\tau)\, z(\tau)\, d\tau \qquad (24.119)$$

$$(t \in [t_0, T]).$$

L'expression (119) montre que la valeur courante de $g(t)$ ne peut se calculer que si l'on connaît le signal d'entrée $z(t)$ dans l'intervalle de temps *futur* $[t, T]$. Par conséquent, pour rendre possible la détermination de $g(t)$ dans l'intervalle de temps $[t_0, T]$, on doit connaître à l'instant même t_0 le signal d'entrée $z(t)$ dans tout l'intervalle de temps *futur* $[t_0, T]$. La nécessité des renseignements préalables sur l'évolution du signal d'entrée $z(t)$ est conditionnée par la forme (82) de la fonctionnelle J_3 à minimiser dans le problème donné.

CHAPITRE 6

SYSTÈMES STOCHASTIQUES

§ 25. Transformation des signaux aléatoires par des systèmes linéaires

Dans de nombreux problèmes le signal appliqué à l'entrée d'un système de commande est la *réalisation* d'un certain *processus aléatoire* dans des conditions où on ne possède aucun renseignement sur cette réalisation, sauf celui relatif aux propriétés statistiques du processus considéré. Ainsi, par exemple, on peut savoir que le processus mis à l'étude est gaussien, on peut connaître sa valeur moyenne et sa fonction de corrélation.

Un tel volume de renseignements sur le signal d'entrée ne suffit pas pour obtenir une information plus complète sur le signal de sortie. Il ne permet que de définir ses propriétés statistiques.

Si le système est régi par des équations différentielles linéaires, la tâche consiste à déterminer les *propriétés statistiques des intégrales de ces équations différentielles* d'après les propriétés statistiques de leurs seconds membres. La résolution de ce problème est donnée dans ce qui suit.

Bien plus grandes sont les difficultés que présente la résolution du problème qui consiste à déterminer les propriétés statistiques des intégrales des équations différentielles dont non seulement les seconds membres, mais encore les coefficients sont des processus aléatoires (fonctions du temps aléatoires).

La résolution des problèmes de ce type pour des équations différentielles non linéaires n'est possible qu'en recourant aux méthodes approchées.

Examinons un système linéaire monovariable dont la fonction de transfert $\Phi(D)$ $\left(\text{où } D = \frac{d}{dt}\right)$ est une fraction régulière et désignons par $k(t)$ la fonction de poids de ce système. Comme nous l'avons montré au § 4, la fonction $p\Phi(p)$ est la transformée de Carson-Heaviside de la fonction $k(t)$:

$$p\Phi(p) \doteqdot k(t), \tag{25.1}$$

soit

$$\Phi(p) = \int_0^\infty k(t)\, e^{-pt}\, dt. \tag{25.2}$$

Pour un système stable, l'intégrale impropre $\int_{-\infty}^{\infty} k(t)\,dt$ est absolument convergente

$$\int_{-\infty}^{\infty} |k(t)|\,dt < \infty,$$

et la transformation de Fourier pour la fonction $k(t)$ s'écrit

$$\Phi(i\omega) = \int_{-\infty}^{\infty} k(t)\,e^{-i\omega t}\,dt. \tag{25.3}$$

Si l'on désigne par $\varphi(t)$ le signal d'entrée, le signal $x(t)$ à la sortie du système, qui à l'instant $t = t_0$ était en équilibre, est égal à

$$x(t) = \int_{t_0}^{t} k(t-\tau)\,\varphi(\tau)\,d\tau = \int_{0}^{t-t_0} k(\xi)\,\varphi(t-\xi)\,d\xi. \tag{25.4}$$

A la limite, pour $t_0 = -\infty$, on a

$$x(t) = \lim_{t_0 \to -\infty} \int_{t_0}^{t} k(t-\tau)\,\varphi(\tau)\,d\tau = \lim_{t_0 \to -\infty} \int_{0}^{t-t_0} k(\xi)\,\varphi(t-\xi)\,d\xi, \tag{25.5}$$

ce que nous allons noter, pour abréger l'écriture,

$$x(t) = \int_{-\infty}^{t} k(t-\tau)\,\varphi(\tau)\,d\tau = \int_{0}^{\infty} k(\xi)\,\varphi(t-\xi)\,d\xi. \tag{25.6}$$

L'expression (6) détermine un *processus stationnaire* dans ce système. Puisque $k(\tau) = 0$ pour $\tau < 0$, l'expression (6) peut se mettre sous la forme

$$x(t) = \int_{-\infty}^{\infty} k(t-\tau)\,\varphi(\tau)\,d\tau = \int_{-\infty}^{\infty} k(\xi)\,\varphi(t-\xi)\,d\xi. \tag{25.7}$$

Dans ce qui suit nous allons supposer que le système étudié est *stable*. Soit un signal d'entrée $\varphi(t)$ produit par un processus aléatoire borné. Etant donné que toute réalisation d'un processus aléatoire borné est une fonction bornée du temps, pour toute réalisation d'un processus aléatoire $\varphi(t)$

$$x(t) = \int_{-\infty}^{\infty} k(\xi)\,\varphi(t-\xi)\,d\xi < \infty. \tag{25.8}$$

L'expression (8) est une relation entre le processus aléatoire $\varphi(t)$ à l'entrée du système et le processus aléatoire $x(t)$ obtenu à sa sortie.

Supposons que le signal d'entrée soit produit par un processus stationnaire aléatoire. L'*espérance mathématique* de ce processus $M[\varphi(t)] = m_\varphi = \text{const}$. Le processus stationnaire à la sortie du système, déterminé par la relation (7), est également un processus aléatoire stationnaire tel que

$$M[x(t)] = \int_{-\infty}^{\infty} k(\xi) M[\varphi(t-\xi)]\, d\xi = m_\varphi \int_{-\infty}^{\infty} k(\xi)\, d\xi. \qquad (25.9)$$

Dans ce qui suit nous allons étudier les processus aléatoires stationnaires $\varphi(t)$ aux moyennes *nulles*.

Par définition, la *fonction de corrélation* du processus aléatoire $x(t)$ défini par l'expression (7) est

$$R_x(t_1, t_2) = M[x(t_1)\, x(t_2)] =$$

$$= M\left[\int_{-\infty}^{\infty} k(\xi)\, \varphi(t_1-\xi)\, d\xi \int_{-\infty}^{\infty} k(\eta)\, \varphi(t_2-\eta)\, d\eta\right]. \qquad (25.10)$$

Les opérations de calcul de l'espérance mathématique et de l'intégration sont commutatives. L'expression (10) peut donc s'écrire

$$R_x(t_1, t_2) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} k(\eta)\, M[\varphi(t_1-\xi)\, \varphi(t_2-\eta)]\, d\eta\, d\xi$$

ou

$$R_x(t_1, t_2) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} k(\eta)\, R_\varphi(t_1-\xi, t_2-\eta)\, d\eta\, d\xi. \qquad (25.11)$$

Puisque $\varphi(t)$ et $x(t)$ sont des processus aléatoires stationnaires, il vient

$$\left.\begin{aligned} R_\varphi(t_1-\xi, t_2-\eta) &= R_\varphi(t_1-t_2+\eta-\xi), \\ R_x(t_1, t_2) &= R_x(t_1-t_2), \end{aligned}\right\} \qquad (25.12)$$

et l'expression (11) devient

$$R_x(t_1-t_2) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} k(\eta)\, R_\varphi(t_1-t_2+\eta-\xi)\, d\eta\, d\xi \qquad (25.13)$$

ou

$$R_x(\tau) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} k(\eta)\, R_\varphi(\tau+\eta-\xi)\, d\eta\, d\xi. \qquad (25.14)$$

Dans ses travaux, A. Khintchine a montré que la densité spectrale $S_f(\omega)$ et la fonction de corrélation $R_f(\tau)$ d'un processus aléatoire sta-

tionnaire $f(t)$ sont liées par les relations

$$\left.\begin{aligned} S_f(\omega) &= \int_{-\infty}^{\infty} R_f(\tau)\, e^{-i\omega\tau}\, d\tau, \\ R_f(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_f(\omega)\, e^{i\omega\tau}\, d\omega. \end{aligned}\right\} \tag{25.15}$$

Conformément à (14) et (15), la *densité spectrale* du processus aléatoire stationnaire $x(t)$ est

$$S_x(\omega) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} k(\eta) \int_{-\infty}^{\infty} R_\varphi(\tau+\eta-\xi)\, e^{-i\omega\tau}\, d\tau\, d\eta\, d\xi. \tag{25.16}$$

En introduisant une nouvelle variable

$$\gamma = \tau + \eta - \xi, \tag{25.17}$$

on ramène l'expression (16) à la forme

$$S_x(\omega) = \int_{-\infty}^{\infty} k(\xi)\, e^{-i\omega\xi}\, d\xi \int_{-\infty}^{\infty} k(\eta)\, e^{i\omega\eta}\, d\eta \int_{-\infty}^{\infty} R_\varphi(\gamma)\, e^{-i\omega\gamma}\, d\gamma$$

ou, en vertu de (3),

$$S_x(\omega) = \Phi(i\omega)\, \Phi(-i\omega) \int_{-\infty}^{\infty} R_\varphi(\gamma)\, e^{-i\omega\gamma}\, d\gamma. \tag{25.18}$$

On tire de (18) et (15) que les densités spectrales des processus aléatoires stationnaires à la sortie et à l'entrée d'un système linéaire stable sont associées par la relation

$$S_x(\omega) = |\Phi(i\omega)|^2\, S_\varphi(\omega). \tag{25.19}$$

Comme $x(t)$ est un processus aléatoire stationnaire à moyenne nulle, sa *variance* s'écrit

$$\mathrm{M}\,[x^2(t)] = R_x(0). \tag{25.20}$$

D'après (15) et (20),

$$\mathrm{M}\,[x^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega)\, d\omega. \tag{25.21}$$

L'expression (19) permet de représenter (21) sous la forme

$$\overline{x^2} = \mathrm{M}\,[x^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi(i\omega)|^2\, S_\varphi(\omega)\, d\omega. \tag{25.22}$$

Pour un processus aléatoire stationnaire, $\zeta(t)$ (à moyenne nulle) du type *bruit blanc*,

$$R_{\zeta}(\tau) = F\delta(\tau), \tag{25.23}$$

où $\delta(\tau)$ est la fonction impulsion de Dirac, et F, l'intensité du bruit. D'après (15), la densité spectrale de ce processus

$$S_{\zeta}(\omega) = F \int_{-\infty}^{\infty} e^{-i\omega\tau}\delta(\tau)\, d\tau = F = \text{const}, \tag{25.24}$$

ce qui peut être adopté comme définition du bruit blanc. La condition (24) traduit le fait que pour *toute* fréquence, la puissance des composantes partielles d'un processus aléatoire $\zeta(t)$ est la même. Le bruit blanc est donc le type le plus intense des perturbations.

Si le signal d'entrée $\varphi(t)$ est un processus aléatoire stationnaire du type bruit blanc à densité spectrale

$$S_{\varphi}(\omega) = F, \tag{25.25}$$

en vertu de (22), la variance du processus aléatoire stationnaire $x(t)$ à la sortie du système est

$$\overline{x^2} = \frac{1}{2\pi} F \int_{-\infty}^{\infty} |\Phi(i\omega)|^2\, d\omega.$$

D'après (20) et (14)

$$\overline{x^2} = R_x(0) = F \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} k(\eta)\,\delta(\eta-\xi)\, d\eta\, d\xi = F \int_{-\infty}^{\infty} [k(\xi)]^2\, d\xi.$$

Ainsi, la variance de $x(t)$ à la sortie du système dans le cas où le signal d'entrée $\varphi(t)$ est un bruit blanc, s'écrit

$$\overline{x^2} = F \int_{-\infty}^{\infty} [k(t)]^2\, dt = \frac{F}{2\pi} \int_{-\infty}^{\infty} |\Phi(i\omega)|^2\, d\omega. \tag{25.26}$$

En introduisant les notations

$$|\Phi(i\omega)| = A(\omega), \tag{25.27}$$

on a

$$\Phi(i\omega) = A(\omega)\, e^{i\psi(\omega)}. \tag{25.28}$$

Ici $A(\omega)$ est la réponse en amplitude et $\psi(\omega)$, la réponse en phase du système.

On tire de (26) et (28) que

$$\overline{x^2} = \frac{F}{2\pi} \int_{-\infty}^{\infty} [A(\omega)]^2\, d\omega. \tag{25.29}$$

Si par Δf on désigne la variable

$$\Delta f=\frac{\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}[A(\omega)]^2\,d\omega}{A_{\max}^2}, \qquad (25.30)$$

l'expression (29) devient

$$\overline{x^2}=FA_{\max}^2\Delta f. \qquad (25.31)$$

La grandeur Δf s'appelle *bande passante* du système.

Cherchons maintenant la fonction de corrélation mutuelle et la densité spectrale mutuelle des processus aléatoires stationnaires $x(t)$ et $\varphi(t)$, où $\varphi(t)$ est le signal d'entrée et $x(t)$ le signal de sortie d'un système linéaire. La fonction de transfert de ce dernier est $\Phi(D)$, où $D=\frac{d}{dt}$.

Par définition, la *fonction de corrélation mutuelle* s'écrit

$$R_{x\varphi}(t_1, t_2)=\mathrm{M}\,[x(t_1)\,\varphi(t_2)]. \qquad (25.32)$$

Puisque d'après (7) pour $t_0\to-\infty$, le signal de sortie $x(t_1)$ est de la forme

$$x(t_1)=\int\limits_{-\infty}^{\infty}k(\xi)\,\varphi(t_1-\xi)\,d\xi, \qquad (25.33)$$

on a

$$R_{x\varphi}(t_1, t_2)=\mathrm{M}\left[\int\limits_{-\infty}^{\infty}k(\xi)\,\varphi(t_1-\xi)\,d\xi\,\varphi(t_2)\right]=$$

$$=\int\limits_{-\infty}^{\infty}k(\xi)\,\mathrm{M}\,[\varphi(t_1-\xi)\,\varphi(t_2)]\,d\xi=\int\limits_{-\infty}^{\infty}k(\xi)\,R_{\varphi}(t_1-t_2-\xi)\,d\xi. \qquad (25.34)$$

Ainsi, pour les processus aléatoires stationnaires:

$$R_{x\varphi}(t_1,\ t_2)=R_{x\varphi}(t_1-t_2)=R_{x\varphi}(\tau), \qquad (25.35)$$

où

$$\tau=t_1-t_2, \qquad (25.36)$$

de plus, en vertu de (34),

$$R_{x\varphi}(\tau)=\int\limits_{-\infty}^{\infty}k(\xi)\,R_{\varphi}(\tau-\xi)\,d\xi. \qquad (25.37)$$

D'une façon analogue

$$R_{\varphi x}(t_1, t_2) = \mathbf{M}[\varphi(t_1)\, x(t_2)] = \mathbf{M}\left[\varphi(t_1) \int_{-\infty}^{\infty} k(\xi)\, \varphi(t_2 - \xi)\, d\xi\right] =$$

$$= \int_{-\infty}^{\infty} k(\xi)\, \mathbf{M}[\varphi(t_1)\, \varphi(t_2 - \xi)]\, d\xi =$$

$$= \int_{-\infty}^{\infty} k(\xi)\, R_{\varphi}(t_1 - t_2 + \xi)\, d\xi. \quad (25.38)$$

Ainsi

$$R_{\varphi x}(\tau) = \int_{-\infty}^{\infty} k(\xi)\, R_{\varphi}(\tau + \xi)\, d\xi, \quad (25.39)$$

où, de même que précédemment (36),

$$\tau = t_1 - t_2.$$

D'après (15) on a la *densité spectrale mutuelle* $S_{x\varphi}(\omega)$

$$S_{x\varphi}(\omega) = \int_{-\infty}^{\infty} R_{x\varphi}(\tau)\, e^{-i\omega\tau}\, d\tau. \quad (25.40)$$

Conformément à (37), l'expression (40) devient

$$S_{x\varphi}(\omega) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} R_{\varphi}(\tau - \xi)\, e^{-i\omega\tau}\, d\tau\, d\xi. \quad (25.41)$$

En introduisant une nouvelle variable

$$\nu = \tau - \xi, \quad (25.42)$$

réduisons l'expression (41) à la forme

$$S_{x\varphi}(\omega) = \int_{-\infty}^{\infty} k(\xi)\, e^{-i\omega\xi}\, d\xi \int_{-\infty}^{\infty} R_{\varphi}(\nu)\, e^{-i\omega\nu}\, d\nu.$$

En vertu de (3) et de (15), cette expression peut s'écrire

$$S_{x\varphi}(\omega) = \Phi(i\omega)\, S_{\varphi}(\omega). \quad (25.43)$$

D'une façon analogue:

$$S_{\varphi x}(\omega) = \int_{-\infty}^{\infty} R_{\varphi x}(\tau)\, e^{-i\omega\tau}\, d\tau =$$

$$= \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} R_{\varphi}(\tau + \xi)\, e^{-i\omega\tau}\, d\tau\, d\xi. \quad (25.44)$$

En introduisant une nouvelle variable

$$\mu = \tau + \xi, \tag{25.45}$$

l'expression (44) se met sous la forme

$$S_{\varphi x}(\omega) = \int_{-\infty}^{\infty} k(\xi)\, e^{i\omega\xi}\, d\xi \int_{-\infty}^{\infty} R_{\varphi}(\mu)\, e^{-i\omega\mu}\, d\mu \tag{25.46}$$

ou, en vertu de (3) et de (15), sous la forme

$$S_{\varphi x}(\omega) = \Phi(-i\omega)\, S_{\varphi}(\omega). \tag{25.47}$$

Etant donné que $S_{\varphi}(\omega) = S_{\varphi}(-\omega)$, les expressions (43) et (47) entraînent

$$S_{\varphi x}(\omega) = S_{x\varphi}(-\omega). \tag{25.48}$$

En retenant que

$$R_{\varphi}(\sigma) = R_{\varphi}(-\sigma),$$

on trouve à partir de (37) que

$$R_{x\varphi}(-\tau) = \int_{-\infty}^{\infty} k(\xi)\, R_{\varphi}(\tau + \xi)\, d\xi,$$

ce qui amène, d'après (39), à la relation

$$R_{\varphi x}(\tau) = R_{x\varphi}(-\tau). \tag{25.49}$$

Cherchons encore la fonction de corrélation mutuelle $R_{x\rho}(\tau)$ des processus aléatoires stationnaires $x(t)$ et $\rho(t)$ (à moyennes nulles), où $x(t)$ est déterminé par l'expression (8).

Par définition

$$R_{x\rho}(t_1, t_2) = \mathbf{M}\,[x(t_1)\rho(t_2)] = \mathbf{M}\left[\int_{-\infty}^{\infty} k(\xi)\, \varphi(t_1 - \xi)\, d\xi\, \rho(t_2)\right] =$$

$$= \int_{-\infty}^{\infty} k(\xi)\, \mathbf{M}\,[\varphi(t_1 - \xi)\, \rho(t_2)]\, d\xi = \int_{-\infty}^{\infty} k(\xi)\, R_{\varphi\rho}(t_1 - t_2 - \xi)\, d\xi. \tag{25.50}$$

En introduisant comme précédemment (36) les notations

$$\tau = t_1 - t_2,$$

on a

$$R_{x\rho}(\tau) = \int_{-\infty}^{\infty} k(\xi)\, R_{\varphi\rho}(\tau - \xi)\, d\xi. \tag{25.51}$$

La densité spectrale mutuelle $S_{x\rho}$ est

$$S_{x\rho}(\omega) = \int_{-\infty}^{\infty} R_{x\rho}(\tau)\, e^{-i\omega\tau}\, d\tau$$

ou, d'après (51),

$$S_{x\rho}(\omega) = \int_{-\infty}^{\infty} k(\xi) \int_{-\infty}^{\infty} R_{\varphi\rho}(\tau-\xi)\, e^{-i\omega\tau}\, d\tau\, d\xi. \tag{25.52}$$

En introduisant de même que précédemment (42) une nouvelle variable

$$\nu = \tau - \xi,$$

on obtient

$$S_{x\rho}(\omega) = \int_{-\infty}^{\infty} k(\xi)\, e^{-i\omega\xi}\, d\xi \int_{-\infty}^{\infty} R_{\varphi\rho}(\nu)\, e^{-i\omega\nu}\, d\nu \tag{25.53}$$

ou

$$S_{x\rho}(\omega) = \Phi(i\omega)\, S_{\varphi\rho}(\omega). \tag{25.54}$$

Si l'on tient compte que d'après (49) et (48)

$$R_{\rho x}(\tau) = R_{x\rho}(-\tau),$$
$$S_{\rho x}(\omega) = S_{x\rho}(-\omega),$$

on a, en vertu de (51) et (54),

$$R_{\rho x}(\tau) = \int_{-\infty}^{\infty} k(\xi)\, R_{\rho\varphi}(\tau+\xi)\, d\xi, \tag{25.55}$$

$$S_{\rho x}(\omega) = \Phi(-i\omega)\, S_{\rho\varphi}(\omega). \tag{25.56}$$

§ 26. Prédiction et filtrage des processus aléatoires unidimensionnels

1. Méthode de A. Kolmogorov et N. Wiener. Processus aléatoires stationnaires. Considérons un système de commande linéaire monovariable où on amène à l'entrée le signal

$$\varphi(t) = h(t) + f(t), \tag{26.1}$$

où $h(t)$ est le signal utile, et $f(t)$, la perturbation. Admettons que $h(t)$ et $f(t)$ sont des processus aléatoires stationnaires à moyennes nulles. Le système a pour but de transformer (et en particulier, reproduire) avec la plus grande précision possible le signal utile $h(t)$.

Désignons par $Y(D)$ la fonction de transfert du système et par $k(t)$, sa fonction de poids.

D'après (25.7) le signal de sortie (sous l'hypothèse que l'instant initial $t_0 \to -\infty$) est de la forme

$$x(t) = \int_{-\infty}^{\infty} \varphi(t-\tau)\, k(\tau)\, d\tau. \tag{26.2}$$

Supposons que le système a pour tâche de prédire le signal d'entrée utile. Pour réaliser la prédiction, imposons au signal de sortie $x(t)$ d'être à l'instant t la meilleure approximation (au sens de minimisation de la variance de l'erreur) par rapport à $h(t+\eta)$, où η = const. Avec $\eta > 0$, on a un problème de prédiction. Avec $\eta = 0$, le problème considéré est un problème de *filtrage* du signal d'entrée [20, 40].

L'erreur d'approximation du signal de sortie $x(t)$ par rapport à $h(t+\eta)$ est

$$\varepsilon(t) = h(t+\eta) - x(t) \tag{26.3}$$

ou, conformément à (2),

$$\varepsilon(t) = h(t+\eta) - \int_{-\infty}^{\infty} \varphi(t-\tau)\, k(\tau)\, d\tau. \tag{26.4}$$

La variance de l'erreur d'approximation s'écrit

$$\mathrm{M}[\varepsilon^2] = \mathrm{M}\left\{\left[h(t+\eta) - \int_{-\infty}^{\infty} \varphi(t-\tau)\, k(\tau)\, d\tau\right]^2\right\}. \tag{26.5}$$

On se propose maintenant de chercher la fonction de poids $k(t)$ qui *minimise* la fonctionnelle $\mathrm{M}[\varepsilon^2]$. Considérons cette *fonction de poids* comme *optimale* et le système auquel elle est associée comme système linéaire optimal.

D'après (5),

$$\mathrm{M}[\varepsilon^2] = \mathrm{M}\left\{h^2(t+\eta) - 2h(t+\eta) \int_{-\infty}^{\infty} \varphi(t-\tau)\, k(\tau)\, d\tau + \right.$$
$$\left. + \left[\int_{-\infty}^{\infty} \varphi(t-\tau)\, k(\tau)\, d\tau\right]^2\right\}$$

ou

$$\mathrm{M}[\varepsilon^2] = \mathrm{M}[h^2(t+\eta)] - 2\int_{-\infty}^{\infty} k(\lambda)\, \mathrm{M}[h(t+\eta)\, \varphi(t-\lambda)]\, d\lambda +$$
$$+ \int_{-\infty}^{\infty} k(\lambda) \int_{-\infty}^{\infty} k(\vartheta)\, \mathrm{M}[\varphi(t-\lambda)\, \varphi(t-\vartheta)]\, d\vartheta\, d\lambda. \tag{26.6}$$

Les processus aléatoires $h(t)$ et $\varphi(t)$ étant des processus stationnaires à moyennes nulles, l'expression (6) devient

$$M[\varepsilon^2] = R_h(0) - 2\int_{-\infty}^{\infty} k(\lambda) R_{h\varphi}(\lambda+\eta)\, d\lambda + \\ + \int_{-\infty}^{\infty} k(\lambda) \int_{-\infty}^{\infty} k(\vartheta) R_\varphi(\lambda-\vartheta)\, d\vartheta\, d\lambda. \quad (26.7)$$

Les fonctions de corrélation $R_h(t)$, $R_\varphi(t)$ et $R_{h\varphi}(t)$ sont supposées connues.

Cherchons à présent la condition nécessaire pour que la fonction $k(t)$ minimise la fonctionnelle

$$I = M[\varepsilon^2]. \quad (26.8)$$

A cet effet, remplaçons $k(t)$ par la fonction $k(t) + \gamma\varkappa(t)$, où γ est un paramètre arbitraire ne dépendant pas de t et $\varkappa(t)$, une fonction arbitraire qui, d'une façon analogue à la fonction $k(t)$, vérifie la contrainte

$$\varkappa(-t) = 0.$$

Introduisons la notation $I + \delta I$ pour la valeur que prend la fonctionnelle I lorsqu'on remplace $k(t)$ par la fonction $k(t) + \gamma\varkappa(t)$. En vertu de (7), on a

$$I + \delta I = R_h(0) - 2\int_{-\infty}^{\infty} [k(\lambda) + \gamma\varkappa(\lambda)] R_{h\varphi}(\lambda+\eta)\, d\lambda + \\ + \int_{-\infty}^{\infty} [k(\lambda) + \gamma\varkappa(\lambda)] \int_{-\infty}^{\infty} [k(\vartheta) + \gamma\varkappa(\vartheta)] R_\varphi(\lambda-\vartheta)\, d\vartheta\, d\lambda. \quad (26.9)$$

Le dernier terme de l'expression (9) peut s'écrire

$$Q = \int_{-\infty}^{\infty} k(\lambda) \int_{-\infty}^{\infty} k(\vartheta) R_\varphi(\lambda-\vartheta)\, d\vartheta\, d\lambda + \\ + \gamma \int_{-\infty}^{\infty} \varkappa(\lambda) \int_{-\infty}^{\infty} k(\vartheta) R_\varphi(\lambda-\vartheta)\, d\vartheta\, d\lambda + \\ + \gamma \int_{-\infty}^{\infty} k(\lambda) \int_{-\infty}^{\infty} \varkappa(\vartheta) R_\varphi(\lambda-\vartheta)\, d\vartheta\, d\lambda + \\ + \gamma^2 \int_{-\infty}^{\infty} \varkappa(\lambda) \int_{-\infty}^{\infty} \varkappa(\vartheta) R_\varphi(\lambda-\vartheta)\, d\vartheta\, d\lambda. \quad (26.10)$$

Les deuxième et troisième termes de l'expression (10) coïncident en raison de la parité de la fonction $R_\varphi(\tau)$.

Désignons par I_2 le facteur affecté à γ^2 du dernier terme de (10)

$$I_2 = \int_{-\infty}^{\infty} \varkappa(\lambda) \int_{-\infty}^{\infty} \varkappa(\vartheta) R_\varphi(\lambda - \vartheta)\, d\vartheta\, d\lambda. \tag{26.11}$$

D'une façon analogue au dernier terme de (7), l'expression (11) peut s'écrire

$$I_2 = \mathrm{M}\left\{\left[\int_{-\infty}^{\infty} \varphi(t-\tau)\varkappa(\tau)\, d\tau\right]^2\right\}. \tag{26.12}$$

Il est évident que

$$I_2 > 0. \tag{26.13}$$

Ainsi, l'expression (9) se ramène à la forme

$$I + \delta I = I - 2\gamma I_1 + \gamma^2 I_2, \tag{26.14}$$

où

$$I_1 = \int_{-\infty}^{\infty} R_{h\varphi}(\lambda + \eta)\varkappa(\lambda)\, d\lambda - \int_{-\infty}^{\infty} \varkappa(\lambda) \int_{-\infty}^{\infty} k(\vartheta) R_\varphi(\lambda - \vartheta)\, d\vartheta\, d\lambda. \tag{26.15}$$

En tenant compte du fait que

$$k(-t) = 0, \quad \varkappa(-t) = 0, \tag{26.16}$$

on peut récrire l'expression (15)

$$I_1 = \int_0^{\infty} \varkappa(\lambda)\left[R_{h\varphi}(\lambda + \eta) - \int_0^{\infty} k(\vartheta) R_\varphi(\lambda - \vartheta)\, d\vartheta\right] d\lambda. \tag{26.17}$$

La condition *nécessaire* pour minimiser la fonctionnelle (7) est

$$\left[\frac{\partial}{\partial\gamma}(I + \delta I)\right]_{\gamma=0} = 0 \text{ pour tout } \varkappa(\lambda). \tag{26.18}$$

D'après (14), la condition (18) devient

$$I_1 = 0. \tag{26.19}$$

On voit sans peine que la condition (19) est non seulement nécessaire, mais encore *suffisante* pour minimiser la fonctionnelle I. A cet effet il faut montrer que quel que soit le choix de la fonction $\varkappa(t)$, on a l'inégalité

$$(I + \delta I)_{I_1=0} > I.$$

L'expression (14) entraîne

$$(I + \delta I)_{I_1=0} = I + \gamma^2 I_2 > I,$$

du fait que d'après (12), $I_2 > 0$.

Ainsi, nous avons démontré que la condition (19) est *nécessaire* et *suffisante* pour minimiser la variance M $[\varepsilon^2]$.

La condition (19) devant être vérifiée pour toute fonction $\varkappa(\lambda)$, conformément à (17) la condition de minimiser la variance M $[\varepsilon^2]$ devient

$$R_{h\varphi}(\lambda+\eta)-\int_0^\infty R_\varphi(\lambda-\vartheta)\,k(\vartheta)\,d\vartheta=0 \qquad (\lambda\geqslant 0). \quad (26.20)$$

De la sorte la fonction de poids $k(t)$ vérifiant l'équation intégrale (20) assure la valeur la plus petite possible de la variance de l'erreur M $[\varepsilon^2]$.

L'équation intégrale (20) a été obtenue par N. Wiener [20]. Pour accorder nos raisonnements avec les notations de Wiener, remplaçons dans (20) la variable d'intégration ϑ par λ, la variable des fonctions de corrélation $R_{h\varphi}$ et R_φ, qui dans (20) se note λ, par τ. L'équation (20) devient alors

$$R_{h\varphi}(\tau+\eta)-\int_0^\infty R_\varphi(\tau-\lambda)\,k(\lambda)\,d\lambda=0 \qquad (\tau\geqslant 0). \quad (26.21)$$

Dans le problème de *filtrage*, c'est-à-dire avec

$$\eta=0, \quad (26.22)$$

en vertu de (21) la fonction de poids optimale $k(t)$ doit satisfaire à l'équation intégrale

$$R_{h\varphi}(\tau)-\int_0^\infty R_\varphi(\tau-\lambda)\,k(\lambda)\,d\lambda=0 \qquad (\tau\geqslant 0). \quad (26.23)$$

2. Résolution de l'équation intégrale définissant la fonction de poids d'un système optimal. Désignons par $S_\varphi(\omega)$ la densité spectrale du signal d'entrée $\varphi(t)$. Puisque $S_\varphi(\omega)$ est une fonction positive paire de ω:

$$S_\varphi(\omega)=S_\varphi(-\omega)>0, \quad (26.24)$$

on peut la mettre sous la forme d'un produit de deux fonctions complexes conjuguées

$$S_\varphi(\omega)=\Psi(i\omega)\,\overline{\Psi(i\omega)}=|\Psi(i\omega)|^2, \quad (26.25)$$

où la fonction $\Psi(p)$ formée à partir de $\Psi(i\omega)$ en remplaçant la variable $i\omega$ par p jouit de la propriété telle que *tous ses zéros et tous ses pôles* se situent *à gauche de l'axe imaginaire* dans le plan de la variable complexe p. La transformation (25) s'appelle *factorisation* de la fonction $S_\varphi(\omega)$. Dans le cas où $S_\varphi(\omega)$ est une fonction rationnelle de ω, la réalisation de la transformation (25) ne présente aucune difficulté.

Vu les propriétés indiquées de la fonction $\Psi(p)$ le système à fonction de transfert $\frac{1}{\Psi(D)}$, où $D = \frac{d}{dt}$, est *stable*.

Imaginons que le système optimal cherché est composé de deux systèmes linéaires reliés en série aux fonctions de transfert respectives $\frac{1}{\Psi(D)}$ et $L(D)$ [76].

La densité spectrale du signal $\sigma(t)$ à la sortie du système à fonction de transfert $1/\Psi(D)$ est, d'après (25.19) et (25),

$$S_\sigma(\omega) = \frac{1}{|\Psi(i\omega)|^2} S_\varphi(\omega) = 1, \qquad (25.26)$$

c'est-à-dire que le signal $\sigma(t)$ est un bruit blanc (processus aléatoire stationnaire) à densité spectrale unitaire. La fonction de corrélation

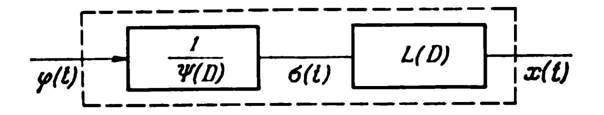


Fig. 26.1

de $\sigma(t)$ est, d'après (25.23) et (26),

$$R_\sigma(\tau) = \delta(\tau), \qquad (26.27)$$

où $\delta(\tau)$ est la fonction impulsion de Dirac.

Désignons par $l(t)$ la fonction de poids du système de la figure 26.1 à fonction de transfert $L(D)$. En vertu de (25.1), on a

$$pL(p) \div l(t). \qquad (26.28)$$

Il s'ensuit du schéma de la figure 26.1 que la fonction de transfert $Y(D)$ du système optimal recherché s'écrit

$$Y(D) = \frac{1}{\Psi(D)} L(D). \qquad (26.29)$$

Pour que $Y(D)$ soit la fonction de transfert d'un système assurant la transformation optimale du signal $\varphi(t)$, il faut que la transformation du signal $\sigma(t)$ par le système à fonction de transfert $L(D)$ soit optimale. A cet effet, la fonction $l(t)$ doit vérifier l'équation intégrale qui, en vertu de (23), s'écrit

$$R_{h\sigma}(\tau) = \int_0^\infty R_\sigma(\tau - \lambda)\, l(\lambda)\, d\lambda \qquad (\tau \geqslant 0). \qquad (26.30)$$

Puisque d'après (27)

$$R_\sigma(\tau - \lambda) = \delta(\tau - \lambda),$$

l'équation (30) devient

$$R_{h\sigma}(\tau)=\int_0^\infty l(\lambda)\,\delta(\tau-\lambda)\,d\lambda,\qquad (\tau\geqslant 0),$$

d'où il résulte

$$l(\tau)=R_{h\sigma}(\tau).\tag{26.31}$$

D'après (25.15),

$$R_{h\sigma}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{h\sigma}(\omega)\,e^{i\omega\tau}\,d\omega.\tag{26.32}$$

D'après (25.56) et compte tenu de fait que $\sigma(t)$ est la sortie et $\varphi(t)$, l'entrée du système à fonction de transfert $\frac{1}{\Psi(D)}$, la densité spectrale $S_{h\sigma}(\omega)$ peut être mise sous la forme

$$S_{h\sigma}(\omega)=\frac{1}{\Psi(-i\omega)}\,S_{h\varphi}(\omega).\tag{26.33}$$

(31), (32) et (33) impliquent que

$$l(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{S_{h\varphi}(\omega)}{\Psi(-i\omega)}\,e^{i\omega\tau}\,d\omega.\tag{26.34}$$

En vertu de (28)

$$L(p)=\int_0^\infty l(\tau)\,e^{-p\tau}\,d\tau,\tag{26.35}$$

d'où, conformément à (34), on a

$$L(p)=\frac{1}{2\pi}\int_0^\infty e^{-p\tau}\,d\tau\int_{-\infty}^{\infty}\frac{S_{h\varphi}(\omega)}{\Psi(-i\omega)}\,e^{i\omega\tau}\,d\omega.\tag{26.36}$$

La fonction de transfert $Y(D)$ $\left(\text{où } D=\frac{d}{dt}\right)$ du système optimal cherché est déterminée, d'après (29) et (36), par l'expression

$$Y(p)=\frac{1}{2\pi\Psi(p)}\int_0^\infty e^{-p\tau}\,d\tau\int_{-\infty}^{\infty}\frac{S_{h\varphi}(\omega)}{\Psi(-i\omega)}\,e^{i\omega\tau}\,d\omega.\tag{26.37}$$

Dans le cas où les densités spectrales $S_\varphi(\omega)$ et $S_{h\varphi}(\omega)$ sont des fonctions rationnelles de ω, la fonction de transfert $Y(D)$ se calcule sans peine. Dans ce cas, d'après (25),

$$S_\varphi(\omega)=\Psi(i\omega)\,\Psi(-i\omega);$$

de plus, comme nous l'avons déjà dit, *tous les zéros et tous les pôles* de la fonction $\Psi(p)$ reposent *à gauche de l'axe imaginaire* dans le plan de la variable complexe p.

Introduisons maintenant les notations

$$K(i\omega)=\frac{S_{h\varphi}(\omega)}{\Psi(-i\omega)} \qquad (26.38)$$

et examinons la fonction $K(p)$ qui se forme à partir de $K(i\omega)$ en remplaçant la variable $i\omega$ par p. Comme $K(p)$ est une fonction rationnelle de p, on peut la développer en une somme de fractions élémentaires et la mettre sous la forme

$$K(p)=K^{+}(p)+K^{-}(p), \qquad (26.39)$$

où tous les pôles de la fonction $K^{+}(p)$ se situent dans le demi-plan gauche de la variable complexe p, et tous les pôles de la fonction $K^{-}(p)$ se trouvent dans le demi-plan droit de p. L'intégrale

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{S_{h\varphi}(\omega)}{\Psi(-i\omega)}e^{i\omega t}\,d\omega$$

est égale à la somme des résidus de la fonction $K(i\zeta)\,e^{i\zeta t}$ par rapport à tous les points singuliers de la fonction $K(i\zeta)$ situés dans le demi-plan supérieur de la variable complexe ζ, c'est-à-dire représente l'original $F(t)$ dont la transformée est la fonction $K^{+}(i\zeta)$.

Etant donné que $K^{+}(i\zeta)$ est la transformée de la fonction $F(t)$, on a

$$\int_{0}^{\infty}F(t)\,e^{-i\omega\tau}\,d\tau=K^{+}(i\omega)\,; \qquad (26.40)$$

l'expression (37) devient donc

$$Y(i\omega)=\frac{K^{+}(i\omega)}{\Psi(i\omega)}, \qquad (26.41)$$

il en résulte que

$$Y(p)=\frac{K^{+}(p)}{\Psi(p)}. \qquad (26.42)$$

L'expression (42) définit la fonction de transfert d'un système optimal dans le cas où les densités spectrales $S_{\varphi}(\omega)$ et $S_{h\varphi}(\omega)$ sont des fonctions rationnelles de ω.

3. Processus aléatoires non stationnaires. Equation intégrale de la fonction de poids optimale. Supposons qu'à l'entrée d'un système linéaire non stationnaire monovariable soit appliqué un processus aléatoire non stationnaire $\zeta(t)$. Désignons par $\vartheta(t,\tau)$ la fonction de poids de ce système. Le signal de sortie du système qui à l'instant

$t = t_0$ était en équilibre ($\xi(t_0) \doteq 0$) sera

$$\xi(t) = \int_{t_0}^{t} \vartheta(t, \tau) \zeta(\tau) \, d\tau. \tag{26.43}$$

Par $\mu(t)$ on désigne la sortie désirée. Si par exemple, $\zeta(t) = m(t) + n(t)$, où $m(t)$ est le signal utile et $n(t)$ les perturbations, la sortie désirée peut être de la forme $\mu(t) = m(t)$ ou présenter une transformation donnée du signal utile $m(t)$.

Introduisons la notation $e(t)$ pour l'erreur de reproduction du signal désiré

$$e(t) = \mu(t) - \xi(t) \tag{26.44}$$

ou, conformément à (43),

$$e(t) = \mu(t) - \int_{t_0}^{t} \vartheta(t, \tau) \zeta(\tau) \, d\tau. \tag{26.45}$$

On a pour la moyenne de $e^2(t)$:

$$I(t) = \mathbf{M}[e^2(t)] = \mathbf{M}\left\{\left[\mu(t) - \int_{t_0}^{t} \vartheta(t, \tau) \zeta(\tau) \, d\tau\right]^2\right\} \tag{26.46}$$

ou

$$I(t) = \mathbf{M}[\mu^2(t)] - 2\mathbf{M}\left[\mu(t) \int_{t_0}^{t} \vartheta(t, \tau) \zeta(\tau) \, d\tau\right] + \\ + \mathbf{M}\left[\int_{t_0}^{t} \vartheta(t, \tau) \zeta(\tau) \, d\tau \int_{t_0}^{t} \vartheta(t, \sigma) \zeta(\sigma) \, d\sigma\right]. \tag{26.47}$$

En admettant que $\mu(t)$ et $\zeta(t)$ sont des processus aléatoires à moyennes nulles et en désignant par

$$\left.\begin{aligned} \varphi_{\mu\mu}(\tau, \sigma) &= \mathbf{M}[\mu(\tau) \mu(\sigma)], \\ \varphi_{\mu\zeta}(\tau, \sigma) &= \mathbf{M}[\mu(\tau) \zeta(\sigma)], \\ \varphi_{\zeta\zeta}(\tau, \sigma) &= \mathbf{M}[\zeta(\tau) \zeta(\sigma)] \end{aligned}\right\} \tag{26.48}$$

les fonctions de corrélation définies sur l'ensemble des fonctions aléatoires respectives, l'expression (47) devient

$$I(t) = \varphi_{\mu\mu}(t, t) - 2 \int_{t_0}^{t} \vartheta(t, \tau) \varphi_{\mu\zeta}(t, \tau) \, d\tau + \\ + \int_{t_0}^{t} \vartheta(t, \tau) \int_{t_0}^{t} \vartheta(t, \sigma) \varphi_{\zeta\zeta}(\tau, \sigma) \, d\sigma \, d\tau. \tag{26.49}$$

La grandeur $I(t)$ qui est la variance de l'erreur dépend de la forme de la fonction de poids, c'est-à-dire représente une fonctionnelle définie sur la classe des fonctions de poids des systèmes linéaires non stationnaires. Si $\vartheta(t, \tau)$ est la fonction de poids optimale, c'est-à-dire si elle minimise la fonctionnelle $I(t)$, cela signifie que pour toute autre fonction de poids $\vartheta(t, \tau) + \gamma h(t, \tau)$ (où γ est un paramètre arbitraire indépendant de t et de τ), la variance de l'erreur est $I(t) + \delta I(t) > I(t)$. En vertu de (49),

$$I(t)+\delta I(t)=\varphi_{\mu\mu}(t, t)-2\int_{t_0}^{t}\vartheta(t, \tau)\varphi_{\mu\zeta}(t, \tau)\,d\tau-$$
$$-2\gamma\int_{t_0}^{t}h(t, \tau)\varphi_{\mu\zeta}(t, \tau)\,d\tau+$$
$$+\int_{t_0}^{t}[\vartheta(t, \tau)+\gamma h(t, \tau)]\int_{t_0}^{t}[\vartheta(t, \sigma)+\gamma h(t, \sigma)]\varphi_{\zeta\zeta}(\tau, \sigma)\,d\sigma\,d\tau. \quad (26.50)$$

Le dernier terme du second membre de (50) s'écrit

$$Q(t)=\int_{t_0}^{t}\vartheta(t, \tau)\int_{t_0}^{t}\vartheta(t, \sigma)\varphi_{\zeta\zeta}(\tau, \sigma)\,d\sigma\,d\tau+$$
$$+\gamma\int_{t_0}^{t}h(t, \tau)\int_{t_0}^{t}\vartheta(t, \sigma)\varphi_{\zeta\zeta}(\tau, \sigma)\,d\sigma\,d\tau+$$
$$+\gamma\int_{t_0}^{t}\vartheta(t, \tau)\int_{t_0}^{t}h(t, \sigma)\varphi_{\zeta\zeta}(\tau, \sigma)\,d\sigma\,d\tau+$$
$$+\gamma^2\int_{t_0}^{t}h(t, \tau)\int_{t_0}^{t}h(t, \sigma)\varphi_{\zeta\zeta}(\tau, \sigma)\,d\sigma\,d\tau. \quad (26.51)$$

Comme

$$\varphi_{\zeta\zeta}(\tau, \sigma) = \varphi_{\zeta\zeta}(\sigma, \tau),$$

l'avant-dernier terme du second membre de (51) devient

$$\gamma\int_{t_0}^{t}\vartheta(t, \tau)\int_{t_0}^{t}h(t, \sigma)\varphi_{\zeta\zeta}(\tau, \sigma)\,d\sigma\,d\tau=$$
$$=\gamma\int_{t_0}^{t}h(t, \sigma)\int_{t_0}^{t}\vartheta(t, \tau)\varphi_{\zeta\zeta}(\sigma, \tau)\,d\tau\,d\sigma. \quad (26.52)$$

Introduisons maintenant les notations

$$I_1(t) = \int_{t_0}^{t} h(t, \sigma)\, \varphi_{\mu\zeta}(t, \sigma)\, d\sigma - \int_{t_0}^{t} h(t, \sigma) \int_{t_0}^{t} \vartheta(t, \tau)\, \varphi_{\zeta\zeta}(\sigma, \tau)\, d\tau\, d\sigma$$

ou

$$I_1(t) = \int_{t_0}^{t} h(t, \sigma) \left[\varphi_{\mu\zeta}(t, \sigma) - \int_{t_0}^{t} \vartheta(t, \tau)\, \varphi_{\zeta\zeta}(\sigma, \tau)\, d\tau\right] d\sigma. \quad (26.53)$$

Désignons par $I_2(t)$ la fonction

$$I_2(t) = \int_{t_0}^{t} h(t, \tau) \int_{t_0}^{t} h(t, \sigma)\, \varphi_{\zeta\zeta}(\tau, \sigma)\, d\sigma\, d\tau. \quad (26.54)$$

L'expression (54) peut s'écrire

$$I_2(t) = \mathrm{M}\left\{\left[\int_{t_0}^{t} h(t, \tau)\, \zeta(\tau)\, d\tau\right]^2\right\},$$

d'où il résulte que

$$I_2(t) > 0.$$

L'expression (50) devient donc

$$I(t) + \delta I(t) = I(t) - 2\gamma I_1(t) + \gamma^2 I_2(t). \quad (26.55)$$

La condition nécessaire pour minimiser la fonctionnelle (46) a pour expression :

$$\frac{\partial}{\partial\gamma}[I(t) + \delta I(t)]_{\gamma=0} = 0 \text{ pour tout } h(t, \tau). \quad (26.56)$$

D'après (55), la condition (56) devient

$$I_1(t) = 0. \quad (26.57)$$

La condition (57) est non seulement nécessaire pour minimiser la fonctionnelle (46), elle est encore suffisante. En effet, pour $I_1(t) = 0$, l'expression (55) s'écrit

$$[I(t) + \delta I(t)]_{I_1(t)=0} = I(t) + \gamma^2 I_2(t) > I(t),$$

du fait que $I_2(t) > 0$ pour tout $h(t, \tau)$.

Comme la condition (57) doit être respectée quelles que soient les fonctions $h(t, \tau)$· d'après (53) la condition de minimisation de $\mathrm{M}[e^2(t)]$ est

$$\int_{t_0}^{t} \vartheta(t, \tau)\, \varphi_{\zeta\zeta}(\sigma, \tau)\, d\tau = \varphi_{\mu\zeta}(t, \sigma), \qquad t_0 \leqslant \sigma < t. \quad (26.58)$$

De la sorte, la fonction de poids optimale $\vartheta(t, \tau)$ doit vérifier l'équation intégrale (58). Cette dernière s'appelle équation de Wiener.

Cherchons maintenant la variance de l'erreur d'un système optimal. D'après (49)

$$M[e^2(t)] = \varphi_{\mu\mu}(t, t) - \int_{t_0}^{t} \vartheta(t, \tau)\, \varphi_{\mu\zeta}(t, \tau)\, d\tau -$$

$$- \int_{t_0}^{t} \vartheta(t, \tau) \left[\varphi_{\mu\zeta}(t, \tau) - \int_{t_0}^{t} \vartheta(t, \sigma)\, \varphi_{\zeta\zeta}(\tau, \sigma)\, d\sigma \right] d\tau. \quad (26.59)$$

Etant donné que la fonction de poids optimale satisfait à la relation (58), pour un système optimal l'expression (59) devient

$$M[e^2(t)] = \varphi_{\mu\mu}(t, t) - \int_{t_0}^{t} \vartheta(t, \tau)\, \varphi_{\mu\zeta}(t, \tau)\, d\tau. \quad (26.60)$$

Voici encore la déduction de deux relations qui, dans le problème exploré, constituent le contenu du lemme sur la projection orthogonale dans un espace de Hilbert, de grand intérêt pour la théorie du filtrage linéaire ([25], p. 286).

L'équation (58) peut s'écrire

$$M[\mu(t)\, \zeta(\sigma)] = \int_{t_0}^{t} \vartheta(t, \tau)\, M[\zeta(\sigma)\, \zeta(\tau)]\, d\tau, \qquad t_0 \leqslant \sigma < t,$$

ou

$$M[\mu(t)\, \zeta(\sigma)] = M\left[\int_{t_0}^{t} \vartheta(t, \tau)\, \zeta(\tau)\, d\tau\, \zeta(\sigma) \right], \qquad t_0 \leqslant \sigma < t.$$

En vertu de (43) cette expression se met sous la forme

$$M[\mu(t)\, \zeta(\sigma)] = M[\xi(t)\, \zeta(\sigma)], \qquad t_0 \leqslant \sigma < t,$$

ou

$$M\{[\mu(t) - \xi(t)]\, \zeta(\sigma)\} = 0, \qquad t_0 \leqslant \sigma < t.$$

Puisque d'après (44) $e(t) = \mu(t) - \xi(t)$, on obtient

$$\varphi_{e\zeta}(t, \sigma) = 0, \qquad t_0 \leqslant \sigma < t. \quad (26.61)$$

Considérons maintenant l'expression (43). En multipliant par $e(t)$ les deux membres de (43), on a

$$e(t)\, \xi(t) = \int_{t_0}^{t} \vartheta(t, \sigma)\, e(t)\, \zeta(\sigma)\, d\sigma.$$

Si l'on remplace les deux membres par leurs espérances mathématiques, il vient

$$\mathrm{M}\,[e(t)\,\xi(t)] = \int_{t_0}^{t} \vartheta(t, \sigma)\,\mathrm{M}\,[e(t)\,\zeta(\sigma)]\,d\sigma.$$

D'après (61), le second membre de cette expression s'annule, et donc,

$$\varphi_{e\xi}(t, t) = 0. \tag{26.62}$$

4. Filtres optimaux de Kalman-Bucy. Etudions le système décrit par l'équation différentielle scalaire

$$\frac{dx}{dt} = f(t)\,x + w(t). \tag{26.63}$$

La présence des perturbations dans les organes de mesure fait que l'état du système se définit avec une erreur de façon que les résultats de l'observation soient de la forme

$$z(t) = x(t) + v(t). \tag{26.64}$$

Supposons que $w(t)$ et $v(t)$ sont des bruits blancs gaussiens non stationnaires à moyennes nulles. Les fonctions de corrélation de ces processus sont

$$\left.\begin{aligned} \varphi_{ww}(t, \tau) &= \mathrm{M}\,[w(t)\,w(\tau)] = q(t)\,\delta(t-\tau), \\ \varphi_{vv}(t, \tau) &= \mathrm{M}\,[v(t)\,v(\tau)] = r(t)\,\delta(t-\tau), \\ \varphi_{wv}(t, \tau) &= \mathrm{M}\,[w(t)\,v(\tau)] = 0, \end{aligned}\right\} \tag{26.65}$$

où $q(t)$ et $r(t)$ sont des fonctions continues dérivables; en outre, $q(t)$ est une fonction non négative et $r(t)$, une fonction positive pour toute valeur de t. Par $\delta(t)$ on désigne la fonction impulsion de Dirac.

Le vecteur $x(t_0)$ est une variable aléatoire gaussienne indépendante de $w(t)$ et $v(t)$, à moyenne nulle $\mathrm{M}\,[x(t_0)] = 0$ et à variance connue.

Pour définir l'estimation $\hat{x}(t)$ de l'état $x(t)$ du système (63), Kalman et Bucy [37] ont proposé d'utiliser un filtre qui est un système régi par l'équation différentielle linéaire inhomogène

$$\frac{d\hat{x}}{dt} = s(t)\,\hat{x} + k(t)\,z(t), \qquad \hat{x}(t_0) = \mathrm{M}\,[x(t_0)] = 0. \tag{26.66}$$

Par ailleurs, les fonctions $s(t)$ et $k(t)$ de l'équation différentielle (66) doivent être choisies de façon à vérifier la condition

$$\mathrm{M}\,\{[x(t) - \hat{x}(t)]^2\} = \min. \tag{26.67}$$

Désignons par $\psi(t, \tau)$ la fonction de poids du système gouverné par l'équation différentielle homogène

$$\frac{d\hat{x}}{dt} = s(t)\,\hat{x}. \tag{26.68}$$

La solution de l'équation différentielle (66) s'écrit alors

$$\hat{x}(t) = \psi(t, t_0)\,\hat{x}(t_0) + \int_{t_0}^{t} \psi(t, \tau)\,k(\tau)\,z(\tau)\,d\tau. \qquad (26.69)$$

Introduisons les notations

$$a(t, \tau) = \psi(t, \tau)\,k(\tau). \qquad (26.70)$$

(70) entraîne que $a(t, t) = \psi(t, t)\,k(t)$, et comme

$$\psi(t, t) = 1,$$

il vient

$$a(t, t) = k(t). \qquad (26.71)$$

Conformément à (69) et (70), pour $\hat{x}(t_0) = 0$, la sortie du système (66) est de la forme

$$\hat{x}(t) = \int_{t_0}^{t} a(t, \tau)\,z(\tau)\,d\tau. \qquad (26.72)$$

Dans le problème de Wiener, l'expression (43) correspond à l'expression (72). Le signal de sortie désirable est la fonction $x(t)$. C'est pourquoi l'équation intégrale de Wiener (58) devient ici

$$\int_{t_0}^{t} a(t, \tau)\,\varphi_{zz}(\sigma, \tau)\,d\tau = \varphi_{xz}(t, \sigma), \qquad t_0 \leqslant \sigma < t. \qquad (26.73)$$

Comme il résulte de la solution du problème de Wiener, la fonction $a(t, \tau)$ qui vérifie l'équation intégrale (73), minimise la fonctionnelle (67).

En dérivant par rapport à t les deux membres de (73), on obtient

$$\int_{t_0}^{t} \frac{\partial}{\partial t} a(t, \tau)\,\varphi_{zz}(\sigma, \tau)\,d\tau + a(t, t)\,\varphi_{zz}(\sigma, t) = \frac{\partial}{\partial t}\varphi_{xz}(t, \sigma), \quad t_0 \leqslant \sigma < t. \qquad (26.74)$$

Les processus aléatoires $x(t)$ et $v(t)$ n'étant pas corrélés et $v(\sigma)$ et $v(t)$ ne l'étant pas pour toute valeur de $\sigma < t$, l'expression

$$\varphi_{zz}(\sigma, t) = \mathrm{M}\,\{[x(\sigma) + v(\sigma)]\,[x(t) + v(t)]\} \qquad (26.75)$$

peut s'écrire

$$\varphi_{zz}(\sigma, t) = \varphi_{xx}(\sigma, t) = \varphi_{zx}(\sigma, t), \quad t_0 \leqslant \sigma < t. \qquad (26.76)$$

Constatons que puisque $\varphi_{xx}(t, \sigma) = \varphi_{xx}(\sigma, t)$, (76) implique la relation

$$\varphi_{zx}(\sigma, t) = \varphi_{xz}(t, \sigma), \qquad t_0 \leqslant \sigma < t. \qquad (26.77)$$

Considérons maintenant le second membre de l'expression (74). (63) conduit à

$$\frac{\partial}{\partial t}\varphi_{xz}(t, \sigma)=\frac{\partial}{\partial t}\mathrm{M}[x(t)\,z(\sigma)]=\mathrm{M}[\dot{x}(t)\,z(\sigma)]=$$
$$=\mathrm{M}[f(t)\,x(t)\,z(\sigma)]+\mathrm{M}[w(t)\,z(\sigma)]. \qquad (26.78)$$

Comme il ressort des équations (63) et (64), le processus aléatoire $z(\sigma)$ n'est corrélé pour aucune valeur de $\sigma < t$ avec le signal d'entrée postérieur $w(t)$, soit

$$\mathrm{M}[w(t)\,z(\sigma)]=0, \qquad t_0 \leqslant \sigma < t, \qquad (26.79)$$

et (78) devient

$$\frac{\partial}{\partial t}\varphi_{xz}(t, \sigma)=f(t)\,\varphi_{xz}(t, \sigma), \qquad t_0 \leqslant \sigma < t. \qquad (26.80)$$

Ainsi, conformément à (76), (77) et (80), l'équation (74) se met sous la forme

$$\int_{t_0}^{t}\frac{\partial}{\partial t}a(t, \tau)\,\varphi_{zz}(\sigma, \tau)\,d\tau+a(t, t)\,\varphi_{xz}(t, \sigma)=f(t)\,\varphi_{xz}(t, \sigma),$$
$$t_0 \leqslant \sigma < t. \qquad (26.81)$$

En remplaçant $\varphi_{xz}(t, \sigma)$ par l'expression (73), on ramène l'expression (81) à la forme

$$\int_{t_0}^{t}\frac{\partial}{\partial t}a(t, \tau)\,\varphi_{zz}(\sigma, \tau)\,d\tau+$$
$$+[a(t, t)-f(t)]\int_{t_0}^{t}a(t, \tau)\,\varphi_{zz}(\sigma, \tau)\,d\tau=0, \qquad t_0 \leqslant \sigma < t, \qquad (26.82)$$

ou

$$\int_{t_0}^{t}\left[f(t)\,a(t, \tau)-a(t, t)\,a(t, \tau)-\frac{\partial}{\partial t}a(t, \tau)\right]\varphi_{zz}(\sigma, \tau)\,d\tau=0,$$
$$t_0 \leqslant \sigma < t. \qquad (26.83)$$

Introduisons la notation

$$b(t, \tau)=f(t)\,a(t, \tau)-a(t, t)\,a(t, \tau)-\frac{\partial}{\partial t}a(t, \tau). \qquad (26.84)$$

'équation (83) peut s'écrire alors

$$\int_{t_0}^{t}b(t, \tau)\,\varphi_{zz}(\sigma, \tau)\,d\tau=0, \qquad t_0 \leqslant \sigma < t. \qquad (26.85)$$

En additionnant les premiers et les seconds membres des équations (73) et (85), on obtient

$$\int_{t_0}^{t} [a(t, \tau) + b(t, \tau)]\, \varphi_{zz}(\sigma, \tau)\, d\tau = \varphi_{xz}(t, \sigma), \qquad t_0 \leqslant \sigma < t. \tag{26.86}$$

Ainsi la fonction $a(t, \tau) + b(t, \tau)$, de même que la fonction $a(t, \tau)$ sont des solutions de l'équation intégrale (73) qui minimise la fonctionnelle (67).

Il est facile de montrer que puisque la fonction $r(t)$ de l'expression (65) est supposée positive pour tout t, la différence de deux solutions de (73) est nulle

$$[a(t, \tau) + b(t, \tau)] - a(t, \tau) = 0, \qquad t_0 \leqslant \tau < t.$$

Pour le démontrer, considérons la relation (61). Dans le problème que nous étudions cette relation devient

$$\mathrm{M}\,[e(t)\, z(\sigma)] = 0, \qquad t_0 \leqslant \sigma < t, \tag{26.87}$$

ou

$$\varphi_{ez}(t, \sigma) = 0, \qquad t_0 \leqslant \sigma < t. \tag{26.88}$$

D'après (72), l'estimation optimale qui correspond à la fonction $a(t, \tau)$ est

$$\hat{x}(t) = \int_{t_0}^{t} a(t, \tau)\, z(\tau)\, d\tau. \tag{26.89}$$

L'estimation optimale associée à la fonction $a(t, \tau) + b(t, \tau)$ s'écrit

$$\check{x}(t) = \int_{t_0}^{t} [a(t, \tau) + b(t, \tau)]\, z(\tau)\, d\tau. \tag{26.90}$$

En désignant par

$$\varkappa(t) = \check{x}(t) - \hat{x}(t), \tag{26.91}$$

on a

$$\varkappa(t) = \int_{t_0}^{t} b(t, \tau)\, z(\tau)\, d\tau. \tag{26.92}$$

La condition (87) entraîne les relations

$$\mathrm{M}\,\{[x(t) - \hat{x}(t)]\, z(\sigma)\} = 0, \quad \mathrm{M}\,\{[x(t) - \check{x}(t)]\, z(\sigma)\} = 0,$$

d'où il s'ensuit que

$$\mathrm{M}\,[\varkappa(t)\, z(\sigma)] = 0. \tag{26.93}$$

Considérons maintenant les relations (89) et (90) dont il résulte que

$$M[\varkappa(t)\hat{x}(t)] = \int_{t_0}^{t} a(t, \sigma) M[\varkappa(t) z(\sigma)] d\sigma,$$

$$M[\varkappa(t)\check{x}(t)] = \int_{t_0}^{t} [a(t, \sigma) + b(t, \sigma)] M[\varkappa(t) z(\sigma)] d\sigma.$$

Puisque d'après (93) les seconds membres de ces relations sont nuls, il vient

$$M[\varkappa(t)\hat{x}(t)] = 0, \quad M[\varkappa(t)\check{x}(t)] = 0.$$

On en déduit que

$$M\{\varkappa(t)[\check{x}(t) - \hat{x}(t)]\} = M[\varkappa(t)\varkappa(t)] = 0. \qquad (26.94)$$

En portant dans (94) la valeur (92) de la fonction $\varkappa(t)$, on obtient

$$\int_{t_0}^{t} b(t, \tau) \int_{t_0}^{t} b(t, \sigma) M[z(\tau) z(\sigma)] d\sigma\, d\tau = 0. \qquad (26.95)$$

Comme

$$M[z(\tau) z(\sigma)] = M[x(\tau) x(\sigma)] + M[v(\tau) v(\sigma)] =$$
$$= M[x(\tau) x(\sigma)] + r(\tau)\delta(\tau - \sigma),$$

l'expression (95) devient

$$\int_{t_0}^{t} b(t, \tau) \int_{t_0}^{t} b(t, \sigma) \{M[x(\tau) x(\sigma)] + r(\tau)\delta(\tau - \sigma)\} d\sigma\, d\tau = 0$$

ou

$$\int_{t_0}^{t} b^2(t, \tau) r(\tau) d\tau + M[\lambda^2(t)] = 0, \qquad (26.96)$$

avec

$$\lambda(t) = \int_{t_0}^{t} b(t, \tau) x(\tau) d\tau.$$

Etant donné que $r(\tau) > 0$, $M[\lambda^2(t)] \geqslant 0$, la relation (96) conduit à

$$b(t, \tau) = 0, \quad t_0 \leqslant \tau < t. \qquad (26.97)$$

Les relations (97) et (84) permettent d'obtenir

$$\frac{\partial}{\partial t} a(t, \tau) = f(t) a(t, \tau) - a(t, t) a(t, \tau). \qquad (26.98)$$

En dérivant par rapport à t les deux membres de l'expression (72), on a

$$\frac{d\hat{x}}{dt} = \int_{t_0}^{t} \frac{\partial}{\partial t} a(t, \tau) z(\tau) d\tau + a(t, t) z(t). \qquad (26.99)$$

On déduit de (99) et (98) que

$$\frac{d\hat{x}}{dt} = \int_{t_0}^{t} [f(t) a(t, \tau) - a(t, t) a(t, \tau)] z(\tau) d\tau + a(t, t) z(t) \qquad (26.100)$$

ou

$$\frac{d\hat{x}}{dt} = f(t) \int_{t_0}^{t} a(t, \tau) z(\tau) d\tau - a(t, t) \int_{t_0}^{t} a(t, \tau) z(\tau) d\tau + a(t, t) z(t). \qquad (26.101)$$

Conformément à (72) et (71), l'équation (101) devient

$$\frac{d\hat{x}}{dt} = [f(t) - k(t)] \hat{x} + k(t) z(t). \qquad (26.102)$$

Les équations (102) et (66) impliquent que la fonction $s(t)$ intervenant dans l'équation (66) s'écrit

$$s(t) = f(t) - k(t). \qquad (26.103)$$

Ainsi le filtre optimal dont l'équation (102) peut s'écrire

$$\frac{d\hat{x}}{dt} = f(t) \hat{x} + k(t) [z(t) - \hat{x}], \qquad (26.104)$$

est un système à retour dans lequel l'écart entre avec un gain $k(t)$.

Il reste encore à définir la forme de la fonction $k(t)$. En portant dans l'équation (102) l'expression (64) de $z(t)$, on obtient

$$\frac{d\hat{x}}{dt} = f(t) \hat{x} + k(t)(x - \hat{x}) + k(t) v(t). \qquad (26.105)$$

Désignons par $e(t)$ l'erreur d'estimation de l'état du système

$$e(t) = x(t) - \hat{x}(t). \qquad (26.106)$$

Les équations (63) et (105) permettent de trouver l'équation différentielle vérifiée par $e(t)$

$$\frac{de}{dt} = [f(t) - k(t)] e + w(t) - k(t) v(t). \qquad (26.107)$$

En vertu de (68), (69) et (103), la fonction de poids de l'équation différentielle homogène

$$\frac{de}{dt} = [f(t) - k(t)] e$$

est $\psi(t, \tau)$. La solution de l'équation (107) s'écrit

$$e(t) = \psi(t, t_0)\, e(t_0) + \int_{t_0}^{t} \psi(t, \tau)\,[w(\tau) - k(\tau)\, v(\tau)]\, d\tau. \qquad (26.108)$$

Cherchons maintenant l'équation différentielle vérifiée par la variance de l'erreur

$$l(t) = \mathrm{M}\,[e^2(t)]. \qquad (26.109)$$

Comme

$$\frac{de^2}{dt} = 2e\frac{de}{dt},$$

d'après (107), on a

$$\frac{de^2(t)}{dt} = 2\,[f(t) - k(t)]\, e^2(t) + 2w(t)\, e(t) - 2k(t)\, v(t)\, e(t). \qquad (26.110)$$

En remplaçant le premier et le second membre de l'équation (110) par leurs espérances mathématiques, on obtient l'équation différentielle par rapport à $l(t)$

$$\frac{dl}{dt} = 2\,[f(t) - k(t)]\, l + 2\mathrm{M}\,[w(t)\, e(t)] - 2k(t)\,\mathrm{M}\,[v(t)\, e(t)]. \qquad (26.111)$$

Pour calculer $\mathrm{M}\,[w(t)\, e(t)]$, considérons l'expression (108). En multipliant ses deux membres par $w(t)$ et en passant aux espérances mathématiques, on obtient

$$\mathrm{M}\,[w(t)\, e(t)] = \psi(t, t_0)\,\mathrm{M}\,[w(t)\, e(t_0)] + \\ + \int_{t_0}^{t} \psi(t, \tau)\,\{\mathrm{M}\,[w(t)\, w(\tau)] - k(\tau)\,\mathrm{M}\,[w(t)\, v(\tau)]\}\, d\tau. \qquad (26.112)$$

D'après (65), $\mathrm{M}\,[w(t)\, v(\tau)] = 0$. Etant donné que l'erreur initiale $e(t_0) = x(t_0) - \hat{x}(t_0)$ n'est pas corrélée avec $w(t)$, il vient: $\mathrm{M}\,[w(t)\, e(t_0)] = 0$. Donc, l'expression (112) se met sous la forme

$$\mathrm{M}\,[w(t)\, e(t)] = \int_{t_0}^{t} \psi(t, \tau)\,\mathrm{M}\,[w(t)\, w(\tau)]\, d\tau$$

ou, d'après (65),

$$\mathrm{M}\,[w(t)\, e(t)] = \int_{t_0}^{t} \psi(t, \tau)\, q(t)\, \delta(t - \tau)\, d\tau = \frac{1}{2}\psi(t, t)\, q(t) = \frac{1}{2} q(t). \qquad (26.113)$$

D'une façon analogue, on a

$$\mathrm{M}\,[v(t)\, e(t)] = -\frac{1}{2} k(t)\, r(t). \qquad (26.114)$$

En vertu de (113) et (114), l'équation différentielle (111) devient

$$\frac{dl}{dt} = 2[f(t) - k(t)]\, l + q(t) + k^2(t)\, r(t). \qquad (26.115)$$

Dans ce qui suit nous aurons besoin des formules (61) et (62) du point précédent. Dans notre cas elles s'écrivent

$$\varphi_{ez}(t, \sigma) = 0, \qquad t_0 \leqslant \sigma < t, \qquad (26.116)$$

$$\varphi_{e\hat{x}}(t, t) = 0. \qquad (26.117)$$

En tenant compte des formules (106) et (64), on a pour la relation (116)

$$M\{[x(t) - \hat{x}(t)][x(\sigma) + v(\sigma)]\} = 0. \qquad (26.118)$$

Il s'ensuit de (63) et (64) que

$$M[x(t)\, v(\sigma)] = 0. \qquad (26.119)$$

En substituant (72) à $\hat{x}(t)$, on obtient

$$M[x(t)\, x(\sigma)] - \int_{t_0}^{t} a(t, \tau)\, M\{[x(\tau) + v(\tau)][x(\sigma) + v(\sigma)]\}\, d\tau = 0$$

ou, en vertu de (119),

$$M[x(t)\, x(\sigma)] - \int_{t_0}^{t} a(t, \tau)\{M[x(\tau)\, x(\sigma)] + M[v(\tau)\, v(\sigma)]\}\, d\tau = 0.$$

D'où, en tenant compte du fait que $M[v(\tau)\, v(\sigma)] = r(\tau)\, \delta(\tau - \sigma)$, on a

$$\varphi_{xx}(t, \sigma) - \int_{t_0}^{t} a(t, \tau)\, \varphi_{xx}(\tau, \sigma)\, d\tau = \int_{t_0}^{t} a(t, \tau)\, r(\tau)\, \delta(\tau - \sigma)\, d\tau,$$
$$t_0 \leqslant \sigma < t,$$

ou

$$\varphi_{xx}(t, \sigma) - \int_{t_0}^{t} a(t, \tau)\, \varphi_{xx}(\tau, \sigma)\, d\tau = a(t, \sigma)\, r(\sigma), \quad t_0 \leqslant \sigma < t. \qquad (26.120)$$

Le premier et le second membre de la relation (120) sont des fonctions continues de σ pour tout t fixé. C'est pourquoi en passant dans (120) à la limite pour $\sigma \to t$ et en retenant que $a(t, t) = k(t)$, on obtient

$$\varphi_{xx}(t, t) - \int_{t_0}^{t} a(t, \tau)\, \varphi_{xx}(\tau, t)\, d\tau = k(t)\, r(t). \qquad (26.121)$$

Puisque d'après (76)

$$\varphi_{xx}(\tau, t) = \varphi_{zx}(\tau, t),$$

l'expression (121) peut être récrite

$$\mathbf{M}\,[x(t)\,x(t)] - \mathbf{M}\left[\int_{t_0}^{t} a(t,\tau)\,z(\tau)\,d\tau\,x(t)\right] = k(t)\,r(t)$$

ou, conformément à (72),

$$\mathbf{M}\,[x(t)\,x(t)] - \mathbf{M}\,[\hat{x}(t)\,x(t)] = \mathbf{M}\,\{[x(t) - \hat{x}(t)]\,x(t)\} = k(t)\,r(t).$$

Cette dernière expression peut s'écrire

$$\mathbf{M}\,\{e(t)\,[e(t) + \hat{x}(t)]\} = k(t)\,r(t). \tag{26.122}$$

Comme d'après (117) $\mathbf{M}\,[e(t)\,\hat{x}(t)] = 0$, on déduit de (122) et (109) que

$$l(t) = k(t)\,r(t). \tag{26.123}$$

Il ressort de (123) que

$$k(t) = \frac{l(t)}{r(t)}. \tag{26.124}$$

Les expressions (115) et (124) donnent la forme définitive de l'équation différentielle vérifiée par la fonction $l(t)$

$$\frac{dl}{dt} = 2f(t)\,l - \frac{1}{r(t)}\,l^2 + q(t). \tag{26.125}$$

L'équation différentielle (125) est une équation de Riccati. Pour la résoudre, il faut donner la valeur initiale $l(t_0)$. Supposons d'après (66) que

$$\hat{x}(t_0) = 0. \tag{26.126}$$

Puisque d'après (106) $e(t_0) = x(t_0) - \hat{x}(t_0)$, conformément à (109) et (126) la valeur initiale $l(t_0) = \mathbf{M}\,[e^2(t_0)]$ devient

$$l(t_0) = \mathbf{M}\,[x(t_0)\,x(t_0)],$$

soit

$$l(t_0) = \varphi_{xx}(t_0, t_0). \tag{26.127}$$

5. Systèmes stationnaires à temps infini d'observation. Examinons le système décrit par les équations différentielles

$$\frac{dx}{dt} = fx + w(t),$$
$$z(t) = x(t) + v(t), \tag{26.128}$$

qui s'obtiennent des équations (63) et (64) pour

$$f(t) = f = \text{const.}$$

Supposons que les processus aléatoires $w\ (t)$ et $v\ (t)$ soient à moyennes nulles et aient comme fonctions de corrélation

$$\begin{aligned} \mathbf{M}\ [w\ (t)\ w\ (\tau)] &= q\delta\ (t-\tau), \\ \mathbf{M}\ [v\ (t)\ v\ (\tau)] &= r\delta\ (t-\tau), \\ \mathbf{M}\ [w\ (t)\ v\ (\tau)] &= 0, \end{aligned} \qquad (26.129)$$

où q et r sont des constantes positives.

D'après (125), l'équation différentielle de Riccati, vérifiée par la variance de l'erreur du filtre optimal $l\ (t)$, s'écrit

$$\frac{dl}{dt} = 2fl - \frac{1}{r}\, l^2 + q. \qquad (26.130)$$

L'équation différentielle (102) du filtre optimal devient

$$\frac{d\hat{x}}{dt} = [f - k\,(t)]\, \hat{x} + k\,(t)\, z\,(t), \qquad (26.131)$$

où, en vertu de (124),

$$k\,(t) = \frac{1}{r}\, l\,(t). \qquad (26.132)$$

Etudions d'abord le cas où l'instant initial de l'intervalle d'observation $t_0 \to -\infty$ et, par conséquent, l'intervalle d'observation $t - t_0$ vérifie la condition $\lim\limits_{t_0 \to -\infty} (t - t_0) = \infty$. Dans ce cas, comme nous allons le montrer dans ce qui suit (155), la solution de l'équation de Riccati (130) a pour $t_0 \to -\infty$ la limite

$$\lim_{t_0 \to -\infty} l\,(t) = \hat{l}, \qquad (26.133)$$

qui existe pour tout t ; de plus, $\hat{l}$ est une constante positive, solution de l'équation algébrique

$$2f\hat{l} - \frac{1}{r}\, \hat{l}^2 + q = 0. \qquad (26.134)$$

La solution positive de l'équation (134) s'écrit

$$\hat{l} = (f + m)\, r, \qquad (26.135)$$

où

$$m = \sqrt{f^2 + \frac{q}{r}} > 0. \qquad (26.136)$$

En outre, d'après (132),

$$k\,(t) = \frac{1}{r}\, \hat{l} = \hat{k}, \qquad (26.137)$$

où

$$\hat{k} = f + m. \tag{26.138}$$

L'équation du filtre optimal (131) pour le cas où $t_0 \to -\infty$ devient

$$\frac{d\hat{x}}{dt} = -m\hat{x} + \hat{k}z(t). \tag{26.139}$$

Le filtre optimal est asymtotiquement stable du fait que $m > 0$.

Considérons maintenant le cas où $t_0 > -\infty$ et cherchons la solution de l'équation différentielle de Riccati (130). A titre d'exemple de la théorie générale énoncée ci-dessous (28.85), montrons que la fonction $l(t)$, qui est solution de (130), peut être exprimée à l'aide de la matrice fondamentale des solutions du système d'équations différentielles linéaires

$$\frac{d\eta_1}{dt} = -f\eta_1 + \frac{1}{r}\eta_2,$$
$$\frac{d\eta_2}{dt} = q\eta_1 + f\eta_2. \tag{26.140}$$

On voit aisément que (140) entraîne

$$\eta_2(t) = l(t)\,\eta_1(t), \tag{26.141}$$

où $l(t)$ est solution de l'équation de Riccati (130). En effet, conformément à (141), la deuxième équation (140) se met sous la forme

$$\frac{dl}{dt}\eta_1(t) + l(t)\frac{d\eta_1}{dt} = q\eta_1(t) + fl(t)\,\eta_1(t). \tag{26.142}$$

En portant dans l'équation (142) au lieu de $\frac{d\eta_1}{dt}$ le second membre de la première équation (140) et en tenant compte de (141), on obtient la relation

$$\left(\frac{dl}{dt} - 2fl + \frac{1}{r}l^2 - q\right)\eta_1(t) = 0. \tag{26.143}$$

Cette égalité est vérifiée identiquement en vertu de l'équation de Riccati (130) qui définit la fonction $l(t)$. Nous avons justifié ainsi la relation (141).

Désignons par $g(t)$ la matrice fondamentale des solutions du système d'équation (140), et par $\Pi(t, t_0)$ la matrice

$$\Pi(t, t_0) = g(t)\,g^{-1}(t_0). \tag{26.144}$$

La solution du système d'équation (140) s'écrit

$$\eta(t) = \Pi(t, t_0)\,\eta(t_0), \tag{26.145}$$

où

$$\Pi(t, t_0) = \begin{bmatrix} \Pi_{11}(t_1, t_0) & \Pi_{12}(t, t_0) \\ \Pi_{21}(t, t_0) & \Pi_{22}(t, t_0) \end{bmatrix}, \quad \eta(t) = \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \end{bmatrix}. \tag{26.146}$$

(145) implique

$$\eta_1(t) = \Pi_{11}(t, t_0)\,\eta_1(t_0) + \Pi_{12}(t, t_0)\,\eta_2(t_0),$$
$$\eta_2(t) = \Pi_{21}(t, t_0)\,\eta_1(t_0) + \Pi_{22}(t, t_0)\,\eta_2(t_0). \qquad (26.147)$$

Puisque d'après (141)

$$\eta_2(t_0) = l(t_0)\,\eta_1(t_0), \qquad (26.148)$$

(147) et (141) entraînent

$$l(t)\,[\Pi_{11}(t, t_0) + \Pi_{12}(t, t_0)\,l(t_0)] =$$
$$= \Pi_{21}(t, t_0) + \Pi_{22}(t, t_0)\,l(t_0), \qquad (26.149)$$

et, par suite,

$$l(t) = \frac{\Pi_{21}(t, t_0) + \Pi_{22}(t, t_0)\,l(t_0)}{\Pi_{11}(t, t_0) + \Pi_{12}(t, t_0)\,l(t_0)}. \qquad (26.150)$$

La matrice fondamentale des solutions du système d'équations (140) s'écrit

$$g(t) = \begin{bmatrix} (m-f)\,e^{mt} & -(m+f)\,e^{-mt} \\ qe^{mt} & qe^{-mt} \end{bmatrix}. \qquad (26.151)$$

Il en résulte que

$$g^{-1}(t_0) = \begin{bmatrix} \frac{1}{2m}\,e^{-mt_0} & \frac{m+f}{2qm}\,e^{-mt_0} \\ -\frac{1}{2m}\,e^{mt_0} & \frac{m-f}{2qm}\,e^{mt_0} \end{bmatrix}. \qquad (26.152)$$

Conformément à (144), (151) et (152),

$$\Pi_{11}(t, t_0) = \frac{m-f}{2m}\,e^{m(t-t_0)} + \frac{m+f}{2m}\,e^{-m(t-t_0)},$$
$$\Pi_{12}(t, t_0) = \frac{m^2-f^2}{2qm}\,e^{m(t-t_0)} - \frac{m^2-f^2}{2qm}\,e^{-m(t-t_0)},$$
$$\Pi_{21}(t, t_0) = \frac{q}{2m}\,e^{m(t-t_0)} - \frac{q}{2m}\,e^{-m(t-t_0)}, \qquad (26.153)$$
$$\Pi_{22}(t, t_0) = \frac{m+f}{2m}\,e^{m(t-t_0)} + \frac{m-f}{2m}\,e^{-m(t-t_0)}.$$

Il s'ensuit de (150) et (153) que

$$l(t) = \frac{n_1(t)}{n_2(t)},$$

$$n_1(t) = \frac{q}{2m}\,e^{m(t-t_0)} - \frac{q}{2m}\,e^{-m(t-t_0)} + \left[\frac{m+f}{2m}\,e^{m(t-t_0)} + \right.$$
$$\left. + \frac{m-f}{2m}\,e^{-m(t-t_0)}\right] l(t_0),$$

$$n_2(t) = \frac{m-f}{2m}\,e^{m(t-t_0)} + \frac{m+f}{2m}\,e^{-m(t-t_0)} +$$
$$+ \left[\frac{m^2-f^2}{2qm}\,e^{m(t-t_0)} - \frac{m^2-f^2}{2qm}\,e^{-m(t-t_0)}\right] l(t_0).$$

En tenant compte que d'après (136) $m^2-f^2=q/r$, la relation précédente se met sous la forme

$$l(t)=\frac{n_1^*(t)}{n_2^*(t)}, \tag{26.154}$$

$$n_1^*(t)=qe^{m(t-t_0)}-qe^{-m(t-t_0)}+[(m+f)\,e^{m(t-t_0)}+(m-f)\,e^{-m(t-t_0)}]\,l(t_0),$$
$$n_2^*(t)=(m-f)\,e^{m(t-t_0)}+(m+f)\,e^{-m(t-t_0)}+$$
$$+\left[\frac{1}{r}e^{m(t-t_0)}-\frac{1}{r}e^{-m(t-t_0)}\right]l(t_0).$$

Remarquons que pour $l(t_0)=0$, $t_0\to-\infty$, l'expression (154) entraîne que

$$\lim_{t_0\to-\infty} l(t)=\frac{q}{m-f}=\frac{q(m+f)}{m^2-f^2}=\frac{q(m+f)}{\frac{q}{r}}=(m+f)\,r=\hat{l}, \tag{26.155}$$

ce qui confirme la validité de la relation (133) donnée dans ce qui précède.

Avec $l(t_0)\neq 0$, $t_0\to-\infty$, (154) conduit à

$$\lim_{t_0\to-\infty} l(t)=\frac{q+(m+f)\,l(t_0)}{m-f+\frac{1}{r}l(t_0)}=\frac{q+\hat{l}\,\frac{1}{r}l(t_0)}{m-f+\frac{1}{r}l(t_0)}. \tag{26.156}$$

D'après (155),

$$m-f=\frac{q}{\hat{l}}.$$

C'est pourquoi

$$m-f+\frac{1}{r}l(t_0)=\frac{q}{\hat{l}}+\frac{1}{r}l(t_0)=\frac{1}{\hat{l}}\left[q+\hat{l}\,\frac{1}{r}l(t_0)\right], \tag{26.157}$$

et l'expression (156) devient

$$\lim_{t_0\to-\infty} l(t)=\hat{l}. \tag{26.158}$$

Soit, maintenant, $t_0>-\infty$. Pour $l(t_0)=0$, on obtient de (154)

$$\lim_{t\to\infty} l(t)=\frac{q}{m-f}=\hat{l}. \tag{26.159}$$

Pour $t_0>-\infty$, $l(t_0)\neq 0$, on a, d'une façon analogue à (156),

$$\lim_{t\to\infty} l(t)=\frac{q+(m+f)\,l(t_0)}{m-f+\frac{1}{r}l(t_0)}=\hat{l}. \tag{26.160}$$

23—0393

Les relations (155), (156), (159) et (160), obtenues ici pour les processus aléatoires unidimensionnels, illustrent les théorèmes généraux de la théorie du filtrage (§ 29) des processus aléatoires multidimensionnels.

6. Systèmes non stationnaires à temps infini d'observation. Revenons maintenant aux équations (63) et (64)

$$\frac{dx}{dt}=f(t)\,x+w(t),$$
$$z(t)=x(t)+v(t).$$

Supposons, comme précédemment, que les processus aléatoires $w(t)$ et $v(t)$ sont à moyennes nulles. Les fonctions de corrélation de ces processus s'écrivent

$$\begin{aligned} \mathbf{M}\,[w(t)\,w(\tau)] &= q(t)\,\delta(t-\tau),\\ \mathbf{M}\,[v(t)\,v(\tau)] &= r(t)\,\delta(t-\tau),\\ \mathbf{M}\,[w(t)\,v(\tau)] &= 0. \end{aligned}$$

Supposons ensuite que les fonctions $f(t)$, $q(t)$ et $r(t)$ vérifient les contraintes

$$\begin{aligned} &|f(t)| \leqslant \beta_1 < \infty,\\ &0<\alpha_2 \leqslant q(t) \leqslant \beta_2 < \infty,\\ &0<\alpha_3 \leqslant r(t) \leqslant \beta_3 < \infty. \end{aligned} \tag{26.161}$$

L'équation différentielle de Riccati, vérifiée par la variance de l'erreur du filtre optimal $l(t)$, est de la forme (125):

$$\frac{dl}{dt}=2f(t)\,l-\frac{1}{r(t)}\,l^2+q(t).$$

Désignons par $\Omega(l)$ le second membre de cette équation

$$\Omega(l)=2f(t)\,l-\frac{1}{r(t)}\,l^2+q(t). \tag{26.162}$$

Pour une valeur fixée $t=t^*$, les zéros de la fonction $\Omega(l)$ sont

$$\begin{aligned} l_1(t^*) &= f(t^*)\,r(t^*)-\sqrt{f^2(t^*)\,r^2(t^*)+q(t^*)\,r(t^*)}<0,\\ l_2(t^*) &= f(t^*)\,r(t^*)+\sqrt{f^2(t^*)\,r^2(t^*)+q(t^*)\,r(t^*)}>0. \end{aligned} \tag{26.163}$$

Les inégalités (161) impliquent que pour tout t^* la valeur $l_2(t^*)$ est comprise entre les limites

$$\alpha_4 \leqslant l_2(t^*) \leqslant \beta_4, \tag{26.164}$$

où

$$\begin{aligned} 0<\alpha_4 &= -\alpha_3\beta_1+\sqrt{\alpha_3^2\beta_1^2+\alpha_2\alpha_3},\\ \beta_4 &= \beta_1\beta_3+\sqrt{\beta_1^2\beta_3^2+\beta_2\beta_3}. \end{aligned} \tag{26.165}$$

Donc, comme on le voit de (162),

$$\frac{dl}{dt} = \Omega(l) > 0 \quad \text{pour} \quad l \in [0, \alpha_4),$$
$$\frac{dl}{dt} = \Omega(l) < 0 \quad \text{pour} \quad l \in (\beta_4, \infty). \tag{26.166}$$

Les inégalités (166) permettent de conclure qu'à partir d'un certain instant $t_1 > t_0$, pour toute condition initiale $l(t_0) \geqslant 0$, la solution de l'équation différentielle de Riccati (125) ne quitte pas le domaine

$$0 < \alpha_4 - \varepsilon \leqslant l(t) \leqslant \beta_4 + \varepsilon < \infty, \tag{26.167}$$

où ε est une petite grandeur correspondante.

Maintenant nous pouvons démontrer la *stabilité asymptotique uniforme* de la solution triviale de l'équation différentielle homogène

$$\frac{d\hat{x}}{dt} = \left[f(t) - \frac{l(t)}{r(t)}\right]\hat{x}, \tag{26.168}$$

construite à partir de l'équation différentielle du filtre optimal (66) avec $z(t) \equiv 0$.

Adoptons comme fonction de Liapounov la fonction

$$V(\hat{x}, t) = \frac{1}{l(t)}\hat{x}^2. \tag{26.169}$$

En raison de (167) et (169), la fonction $V(\hat{x}, t)$ satisfait aux conditions

$$\frac{1}{\beta_4+\varepsilon}\hat{x}^2 \leqslant V(\hat{x}, t) \leqslant \frac{1}{\alpha_4-\varepsilon}\hat{x}^2 \quad \text{pour} \quad t > t_1. \tag{26.170}$$

La dérivée de la fonction $V(\hat{x}, t)$ par rapport au temps est de la forme

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \hat{x}}\frac{d\hat{x}}{dt} = -\frac{1}{l^2(t)}\frac{dl}{dt}\hat{x}^2 + \frac{1}{l(t)}2\hat{x}\frac{d\hat{x}}{dt}$$

ou

$$\frac{dV}{dt} = -\left[\frac{1}{l(t)}\frac{dl}{dt} - 2\frac{d\hat{x}}{dt}\frac{1}{\hat{x}}\right]V. \tag{26.171}$$

En portant la valeur de $\frac{dl}{dt}$ tirée de (125) et la valeur de $\frac{d\hat{x}}{dt}$ tirée de (168), réduisons l'expression (171) à la forme

$$\frac{dV}{dt} = -\left[\frac{q(t)}{l(t)} + \frac{l(t)}{r(t)}\right]V \leqslant -\left(\frac{\alpha^2}{\beta_4+\varepsilon} + \frac{\alpha_4-\varepsilon}{\beta_3}\right)V \quad \text{pour } t \geqslant t_1. \tag{26.172}$$

Ainsi, $\frac{dV}{dt}$ est strictement négative pour tout $t \geqslant t_1$ et, donc, le *filtre optimal est uniformément asymptotiquement stable.*

Les résultats obtenus permettent également d'établir certaines propriétés des solutions de l'équation de Riccati (125). Désignons par $l^{(a)}(t)$ et $l^{(b)}(t)$ les solutions de l'équation différentielle (125) sous des conditions initiales $l^{(a)}(t_0) = l_0^{(a)}$ et $l^{(b)}(t_0) = l_0^{(b)}$ respectivement. Ici $l_0^{(a)}$ et $l_0^{(b)}$ sont des grandeurs positives (ou nulles).

En introduisant les notations

$$\delta l(t) = l^{(a)}(t) - l^{(b)}(t), \tag{26.173}$$

on aura

$$\frac{d}{dt}[\delta l(t)] = \frac{dl^{(a)}(t)}{dt} - \frac{dl^{(b)}(t)}{dt}, \tag{26.174}$$

ou, d'après (125),

$$\frac{d}{dt}[\delta l(t)] = 2f(t)\, l^{(a)}(t) - \frac{1}{r(t)}[l^{(a)}(t)]^2 + \\ + q(t) - \left\{ 2f(t)\, l^{(b)}(t) - \frac{1}{r(t)}[l^{(b)}(t)]^2 + q(t) \right\}. \tag{26.175}$$

En désignant conformément à (103) et (124)

$$s^{(a)}(t) = f(t) - \frac{1}{r(t)} l^{(a)}(t), \quad s^{(b)}(t) = f(t) - \frac{1}{r(t)} l^{(b)}(t), \tag{26.176}$$

on met l'équation (175) sous la forme

$$\frac{d}{dt}[\delta l(t)] = [s^{(a)}(t) + s^{(b)}(t)]\, \delta l(t). \tag{26.177}$$

Désignons conformément à (68) et (69) par $\psi^{(a)}(t, \tau)$ la fonction de poids de l'équation différentielle

$$\frac{d\xi^{(a)}}{dt} = s^{(a)}(t)\, \xi^{(a)}, \tag{26.178}$$

et par $\psi^{(b)}(t, \tau)$ la fonction de poids de l'équation différentielle

$$\frac{d\xi^{(b)}}{dt} = s^{(b)}(t)\, \xi^{(b)}. \tag{26.179}$$

Les solutions de ces équations sont

$$\xi^{(a)}(t) = \psi^{(a)}(t, t_0)\, \xi^{(a)}(t_0), \quad \xi^{(b)}(t) = \psi^{(b)}(t, t_0)\, \xi^{(b)}(t_0). \tag{26.180}$$

Puisque

$$\begin{gathered} \frac{d}{dt}\psi^{(a)}(t, t_0) = s^{(a)}(t)\, \psi^{(a)}(t, t_0), \\ \frac{d}{dt}\psi^{(b)}(t, t_0) = s^{(b)}(t)\, \psi^{(b)}(t, t_0), \end{gathered} \tag{26.181}$$

on montre sans peine que la solution de l'équation différentielle (177) s'écrit

$$\delta l(t) = \psi^{(a)}(t, t_0)\, \psi^{(b)}(t, t_0)\, \delta l(t_0). \tag{26.182}$$

D'après (104) et (124), pour la valeur initiale $l^{(a)}(t_0) = l_0^{(a)}$ de la variance de l'erreur, le filtre optimal est régi par l'équation différentielle

$$\frac{d\hat{x}^{(a)}}{dt} = s^{(a)}(t)\,\hat{x}^{(a)} + \frac{l^{(a)}(t)}{r(t)}\,z(t), \qquad (26.183)$$

et pour la valeur initiale $l^{(b)}(t_0) = l_0^{(b)}$ de la variance de l'erreur, l'équation du filtre optimal est

$$\frac{d\hat{x}^{(b)}}{dt} = s^{(b)}(t)\,\hat{x}^{(b)} + \frac{l^{(b)}(t)}{r(t)}\,z(t). \qquad (26.184)$$

Nous avons montré précédemment que quelle que soit la valeur initiale de la variance de l'erreur $l(t)$, le filtre optimal est asymptotiquement stable. Cela signifie que

$$\lim_{t\to\infty} \psi^{(a)}(t, t_0) = 0, \qquad \lim_{t\to\infty} \psi^{(b)}(t, t_0) = 0. \qquad (26.185)$$

Il résulte de (182) et (185) que $\lim\limits_{t\to\infty} [\delta l(t)] = 0$ ou

$$\lim_{t\to\infty} [l^{(a)}(t) - l^{(b)}(t)] = 0, \qquad (26.186)$$

c'est-à-dire que *les solutions de l'équation de Riccati* (125), *associées à des conditions initiales différentes, se rapprochent asymptotiquement quand* $t \to \infty$. Cela signifie que toute solution de (125) tend vers une solution particulière $\tilde{l}(t)$ contenue dans le domaine (167). Désignant par $l(t; l_0, t_0)$ la solution de (125) qui satisfait à la condition initiale $l(t_0) = l_0$, on peut définir $\tilde{l}(t)$ de la façon suivante

$$\lim_{t_0\to-\infty} l(t;\ 0,\ t_0) = \tilde{l}(t). \qquad (26.187)$$

De l'exposé ci-dessus il s'ensuit que toute solution $l(t; l_0, t_0)$ de l'équation de Riccati (125) qui vérifie la condition initiale $l(t_0) = l_0 \geqslant 0$, où $t_0 > -\infty$, converge uniformément vers $\tilde{l}(t)$, où la fonction $\tilde{l}(t)$ est définie par l'expression (187). En ce sens, $\tilde{l}(t)$ est un état d'équilibre mobile qui vérifie l'équation différentielle (125).

§ 27. Filtrage des processus aléatoires multidimensionnels

1. Equation intégrale matricielle de Wiener de la fonction de poids optimale. Examinons le problème de filtrage des processus aléatoires multidimensionnels. Bornons-nous à l'étude des filtres constitués par des systèmes régis par des équations différentielles

Tableau

	Équation différentielle	Matrice fondamentale	Matrice de Cauchy
Système initial	$\frac{dx}{dt}=A(t)\,x$	$\theta(t)$	$\Phi(t,\tau)=$ $=\theta(t)\,\theta^{-1}(\tau)$
Système conjugué du système initial	$\frac{dy}{dt}=-A^*(t)\,y$	$\nu(t)$	$h(t,\tau)=$ $=\nu(t)\,\nu^{-1}(\tau)$
Filtre optimal	$\frac{d\rho}{dt}=F(t)\,\rho,$ $F(t)=A(t)-G(t)\,C(t),$ $G(t)=\Sigma(t)\,C^*(t)\,R^{-1}(t)$	$\varkappa(t)$	$\Psi(t,\tau)=$ $=\varkappa(t)\,\varkappa^{-1}(\tau)$
Système conjugué du filtre optimal	$\frac{d\tilde{\xi}}{dt}=-F^*(t)\,\tilde{\xi},$ $F^*(t)=A^*(t)-C^*(t)\,G^*(t),$ $G^*(t)=R^{-1}(t)\,C(t)\,\Sigma(t)$	$\tilde{\vartheta}(t)$	$\Lambda(t,\tau)=$ $=\tilde{\vartheta}(t)\,\tilde{\vartheta}^{-1}(\tau)$
Filtre du problème d'estimation	$\frac{d\hat{x}}{dt}=[A(t)-$ $-\mathcal{M}^{-1}(t,t_0)\,C^*(t)\times$ $\times R^{-1}(t)\,C(t)]\,\hat{x}$	$\vartheta(t)$	$\mathcal{T}(t,\tau)=$ $=\vartheta(t)\,\vartheta^{-1}(\tau)$
Système d'équations différentielles vectorielles qui sert pour résoudre l'équation de Riccati	$\frac{d\xi}{dt}=-A^*(t)\,\xi+$ $+C^*(t)\,R^{-1}(t)\,C(t)\,\mu,$ $\frac{d\mu}{dt}=B(t)\,Q(t)\,B^*(t)\,\xi+$ $+A(t)\,\mu$	$g(t)$	$\Pi(t,\tau)=$ $=g(t)\,g^{-1}(\tau)$

linéaires à coefficients variables

$$\frac{d\rho}{dt}=F(t)\,\rho+G(t)\,z(t). \qquad (27.1)$$

Ici ρ est le vecteur de dimension n; $F(t)$, la matrice $n\times n$; $z(t)$, le vecteur de dimension m; $G(t)$, la matrice $n\times m$ *.

* Les notations adoptées pour les matrices fondamentales des solutions et les matrices de Cauchy des équations différentielles vectorielles et des systèmes d'équations vectorielles étudiés aux §§ 27-29 sont données dans le tableau.

Le processus aléatoire non stationnaire $z(t)$ est un signal fourni à l'entrée du système. Le signal de sortie s'écrit

$$\rho(t) = \Psi(t, t_0)\,\rho(t_0) + \int_{t_0}^{t} \Psi(t, \tau)\,G(\tau)\,z(\tau)\,d\tau, \qquad (27.2)$$

où

$$\Psi(t, \tau) = \varkappa(t)\,\varkappa^{-1}(\tau), \qquad (27.3)$$

et où $\varkappa(t)$ désigne la matrice fondamentale des solutions de l'équation différentielle homogène

$$\frac{d\rho}{dt} = F(t)\,\rho. \qquad (27.4)$$

Avec

$$\rho(t_0) = 0, \qquad (27.5)$$

la sortie du système devient

$$\rho(t) = \int_{t_0}^{t} \Psi(t, \tau)\,G(\tau)\,z(\tau)\,d\tau$$

ou

$$\rho(t) = \int_{t_0}^{t} a(t, \tau)\,z(\tau)\,d\tau, \qquad (27.6)$$

où $a(t, \tau)$ désigne la matrice $n \times m$

$$a(t, \tau) = \Psi(t, \tau)\,G(\tau). \qquad (27.7)$$

Introduisons la notation $x(t)$ pour le signal vectoriel de dimension n qu'on voudrait obtenir à la sortie du système.

Par $e(t)$ désignons l'erreur de reproduction du signal à obtenir

$$e(t) = x(t) - \rho(t), \qquad (27.8)$$

ou, en vertu de (6),

$$e(t) = x(t) - \int_{t_0}^{t} a(t, \tau)\,z(\tau)\,d\tau. \qquad (27.9)$$

Supposons que $x(t)$ et $z(t)$ sont des processus aléatoires à moyennes nulles. De plus, conformément à (5), $\rho(t)$ est également un processus aléatoire à moyenne nulle.

Les matrices de corrélation des processus aléatoires concernés *

* Ici, de même que précédemment, l'astérisque marque la matrice transposée. Le produit scalaire des vecteurs a et b est noté $\langle a, b\rangle$. Le symbole Sp A désigne la trace de la matrice A.

sont

$$\begin{aligned}\operatorname{cov}[x(\tau);\ x(\sigma)] &= \mathbf{M}[x(\tau)\,x^*(\sigma)],\\ \operatorname{cov}[x(\tau);\ z(\sigma)] &= \mathbf{M}[x(\tau)\,z^*(\sigma)],\\ \operatorname{cov}[z(\tau);\ z(\sigma)] &= \mathbf{M}[z(\tau)\,z^*(\sigma)].\end{aligned} \tag{27.10}$$

Adoptons comme mesure de la performance du filtre (I) la fonctionnelle

$$J(t) = \mathbf{M}[\langle e(t),\ e(t)\rangle] = \mathbf{M}[e^*(t)\,e(t)] = \operatorname{Sp}\{\mathbf{M}[e(t)\,e^*(t)]\}. \tag{27.11}$$

Conformément à (9),

$$\begin{aligned}\mathbf{M}[e(t)\,e^*(t)] = \mathbf{M}\Big\{\Big[x(t) - \int_{t_0}^{t} a(t,\ \tau)\,z(\tau)\,d\tau\Big]\Big[x(t) -{}\\ - \int_{t_0}^{t} a(t,\ \sigma)\,z(\sigma)\,d\sigma\Big]^*\Big\} = \mathbf{M}[x(t)\,x^*(t)] -{}\\ - \int_{t_0}^{t} \mathbf{M}[x(t)\,z^*(\sigma)]\,a^*(t,\ \sigma)\,d\sigma -{}\\ - \int_{t_0}^{t} a(t,\ \tau)\,\mathbf{M}[z(\tau)\,x^*(t)]\,d\tau +{}\\ + \int_{t_0}^{t} a(t,\ \tau) \int_{t_0}^{t} \mathbf{M}[z(\tau)\,z^*(\sigma)]\,a^*(t,\ \sigma)\,d\sigma\,d\tau.\end{aligned} \tag{27.12}$$

On voit aisément que les matrices $n \times n$

$$\begin{aligned}L_1 &= \int_{t_0}^{t} \mathbf{M}[x(t)\,z^*(\sigma)]\,a^*(t,\ \sigma)\,d\sigma,\\ L_2 &= \int_{t_0}^{t} a(t,\ \tau)\,\mathbf{M}[z(\tau)\,x^*(t)]\,d\tau\end{aligned} \tag{27.13}$$

satisfont à la condition

$$L_1 = L_2^*, \tag{27.14}$$

d'où l'on déduit que

$$\operatorname{Sp}(L_1 + L_2) = 2\operatorname{Sp} L_1. \tag{27.15}$$

Ainsi, la fonctionnelle (11) devient

$$\begin{aligned}J(t) = \operatorname{Sp}\Big\{\mathbf{M}[x(t)\,x^*(t)] - 2\int_{t_0}^{t} \mathbf{M}[x(t)\,z^*(\sigma)]\,a^*(t,\ \sigma)\,d\sigma +{}\\ + \int_{t_0}^{t} a(t,\ \tau) \int_{t_0}^{t} \mathbf{M}[z(\tau)\,z^*(\sigma)]\,a^*(t,\ \sigma)\,d\sigma\,d\tau\Big\}.\end{aligned} \tag{27.16}$$

La grandeur $J(t)$, mesure de l'erreur du filtre, dépend de la forme de la fonction de poids $a(t, \tau) = \Psi(t, \tau) G(\tau)$. Si $a(t, \tau)$ est une fonction de poids optimale, c'est-à-dire si elle minimise la fonctionnelle $J(t)$, cela signifie que pour toute autre fonction de poids $a(t, \tau) + \gamma\tilde{h}(t, \tau)$ (où γ est un paramètre scalaire arbitraire indépendant de t et de τ), la mesure de l'erreur du filtre est $J(t) + \delta J(t) > J(t)$.

En vertu de (16), on a

$$J(t)+\delta J(t)=\mathrm{Sp}\Big\{\mathrm{M}[x(t)x^*(t)]-2\int_{t_0}^{t}\mathrm{M}[x(t)z^*(\sigma)][a(t,\sigma)+\gamma\tilde{h}(t,\sigma)]^*\,d\sigma+\int_{t_0}^{t}[a(t,\tau)+\gamma\tilde{h}(t,\tau)]\int_{t_0}^{t}\mathrm{M}[z(\tau)z^*(\sigma)][a(t,\sigma)+\gamma\tilde{h}(t,\sigma)]^*\,d\sigma\,d\tau\Big\}. \quad (27.17)$$

L'expression (17) peut être ramenée à la forme

$$J(t)+\delta J(t)=J(t)+\mathrm{Sp}\Big\{-2\gamma\int_{t_0}^{t}\mathrm{M}[x(t)z^*(\sigma)]\tilde{h}^*(t,\sigma)\,d\sigma+\gamma\int_{t_0}^{t}\tilde{h}(t,\tau)\int_{t_0}^{t}\mathrm{M}[z(\tau)z^*(\sigma)]a^*(t,\sigma)\,d\sigma\,d\tau+\gamma\int_{t_0}^{t}a(t,\tau)\int_{t_0}^{t}\mathrm{M}[z(\tau)z^*(\sigma)]\tilde{h}^*(t,\sigma)\,d\sigma\,d\tau+\gamma^2\int_{t_0}^{t}\tilde{h}(t,\tau)\int_{t_0}^{t}\mathrm{M}[z(\tau)z^*(\sigma)]\tilde{h}^*(t,\sigma)\,d\sigma\,d\tau. \quad (27.18)$$

En introduisant les notations

$$S_1=\int_{t_0}^{t}\tilde{h}(t,\tau)\int_{t_0}^{t}\mathrm{M}[z(\tau)z^*(\sigma)]a^*(t,\sigma)\,d\sigma\,d\tau, \quad (27.19)$$

$$S_2=\int_{t_0}^{t}a(t,\tau)\int_{t_0}^{t}\mathrm{M}[z(\tau)z^*(\sigma)]\tilde{h}^*(t,\sigma)\,d\sigma\,d\tau,$$

$$l(t)=\int_{t_0}^{t}\tilde{h}(t,\tau)z(\tau)\,d\tau \quad (27.20)$$

et en tenant compte de (6), on obtient

$$S_1 = \mathrm{M}\int_{t_0}^{t} \tilde{h}(t, \tau) z(\tau)\, d\tau \int_{t_0}^{t} [a(t, \sigma) z(\sigma)]^* d\sigma = \mathrm{M}[l(t)\rho^*(t)],$$

$$S_2 = \mathrm{M}\int_{t_0}^{t} a(t, \tau) z(\tau)\, d\tau \int_{t_0}^{t} [\tilde{h}(t, \sigma) z(\sigma)]^* d\sigma = \mathrm{M}[\rho(t) l^*(t)].$$

Donc

$$S_2 = S_1^* \tag{27.21}$$

et par suite,

$$\operatorname{Sp} S_1 = \operatorname{Sp} S_2. \tag{27.22}$$

Si l'on tient compte de (22), il vient

$$\operatorname{Sp}\left[-2\gamma\int_{t_0}^{t}\mathrm{M}[x(t) z^*(\sigma)]\tilde{h}^*(t, \sigma)\, d\sigma + \gamma S_1 + \gamma S_2\right] =$$

$$= \operatorname{Sp}\left[-2\gamma\int_{t_0}^{t}\mathrm{M}[x(t) z^*(\sigma)]\tilde{h}^*(t, \sigma)\, d\sigma + \right.$$

$$\left. + 2\gamma\int_{t_0}^{t} a(t, \tau)\int_{t_0}^{t}\mathrm{M}[z(\tau) z^*(\sigma)]\tilde{h}^*(t, \sigma)\, d\sigma\, d\tau\right] =$$

$$= -2\gamma \operatorname{Sp}\left[\int_{t_0}^{t}\left\{\mathrm{M}[x(t) z^*(\sigma)] - \right.\right.$$

$$\left.\left. - \int_{t_0}^{t} a(t, \tau)\mathrm{M}[z(\tau) z^*(\sigma)]\, d\tau\right\}\tilde{h}^*(t, \sigma)\, d\sigma\right] = -2\gamma \operatorname{Sp} J_1(t), \tag{27.23}$$

où

$$J_1(t) = \int_{t_0}^{t}\left\{\mathrm{M}[x(t) z^*(\sigma)] - \int_{t_0}^{t} a(t, \tau)\mathrm{M}[z(\tau) z^*(\sigma)]\, d\tau\right\}\tilde{h}^*(t, \sigma)\, d\sigma. \tag{27.24}$$

Considérons maintenant le dernier terme de l'expression (18). En introduisant la notation

$$J_2(t) = \int_{t_0}^{t}\tilde{h}(t, \tau)\int_{t_0}^{t}\mathrm{M}[z(\tau) z^*(\sigma)]\tilde{h}^*(t, \sigma)\, d\sigma\, d\tau, \tag{27.25}$$

on a en vertu de (20)

$$J_2(t) = \mathrm{M}\int_{t_0}^{t}\tilde{h}(t, \tau)\int_{t_0}^{t} z(\tau) z^*(\sigma)\tilde{h}^*(t, \sigma)\, d\sigma\, d\tau =$$

$$= \mathrm{M}\left\{\int_{t_0}^{t}\tilde{h}(t, \tau) z(\tau)\, d\tau\int_{t_0}^{t}[\tilde{h}(t, \sigma) z(\sigma)]^* d\sigma\right\} = \mathrm{M}[l(t) l^*(t)]. \tag{27.26}$$

On en déduit que

$$\operatorname{Sp} J_2(t) = \mathrm{M}\,[l^*(t)\, l(t)] = \mathrm{M}\,[\langle l(t),\ l(t)\rangle] \tag{27.27}$$

et, donc,

$$\operatorname{Sp} J_2(t) > 0. \tag{27.28}$$

D'après (23) et (25), l'expression (17) devient

$$J(t) + \delta J(t) = J(t) - 2\gamma \operatorname{Sp} J_1(t) + \gamma^2 \operatorname{Sp} J_2(t), \tag{27.29}$$

où $J_1(t)$ et $J_2(t)$ sont définis par les expressions (24) et (25).

La condition nécessaire de minimisation de la fonctionnelle (11) s'écrit

$$\frac{\partial}{\partial\gamma}[J(t)+\delta J(t)]_{\gamma=0}=0 \quad \text{pour tout } \tilde{h}(t,\tau). \tag{27.30}$$

En raison de (29), la condition (30) prend la forme

$$\operatorname{Sp} J_1(t) = 0. \tag{27.31}$$

La condition (31) est non seulement nécessaire mais encore suffisante pour minimiser la fonctionnelle (11). En effet, pour $\operatorname{Sp} J_1(t) = 0$, l'expression (29) devient

$$[J(t)+\delta J(t)]_{J_1(t)=0} = J(t)+\gamma^2 \operatorname{Sp} J_2(t) > J(t) \tag{27.32}$$

du fait que d'après (27) $\operatorname{Sp} J_2(t) > 0$ pour tout $\tilde{h}(t,\tau)$.

Puisque la condition (31) doit être remplie quelles que soient les fonctions $\tilde{h}(t,\tau)$, en vertu de (24) la condition de minimisation de la fonctionnelle $\mathrm{M}\,[\langle e(t),\ e(t)\rangle]$ se met sous la forme

$$\mathrm{M}\,[x(t)\, z^*(\sigma)] = \int_{t_0}^{t} a(t,\tau)\, \mathrm{M}\,[z(\tau)\, z^*(\sigma)]\, d\tau, \quad t_0 \leqslant \sigma < t. \tag{27.33}$$

De la sorte, la matrice $a(t,\tau)$ de type $n \times m$, qui est une fonction de poids optimale du filtre, doit satisfaire à l'équation intégrale matricielle (33). Cette équation s'appelle équation de Wiener.

L'équation intégrale de Wiener (33) entraîne les deux relations suivantes qui traduisent le contenu du lemme utilisé dans la théorie du filtrage linéaire [25, p. 238], relatif à la projection orthogonale dans un espace de Hilbert.

L'équation (33) peut s'écrire

$$\mathrm{M}\left\{\left[x(t) - \int_{t_0}^{t} a(t,\tau)\, z(\tau)\, d\tau\right] z^*(\sigma)\right\} = 0, \quad t_0 \leqslant \sigma < t,$$

ou, d'après (6),

$$\mathrm{M}\,\{[x(t) - \rho(t)]\, z^*(\sigma)\} = 0, \qquad t_0 \leqslant \sigma < t.$$

Conformément à (8) cette relation prend la forme

$$\mathrm{M}\,[e(t)\, z^*(\sigma)] = 0, \qquad t_0 \leqslant \sigma < t. \tag{27.34}$$

Pour obtenir la deuxième relation, considérons l'expression (6). En transposant ses premier et second membres, on obtient

$$\rho^* (t) = \int_{t_0}^{t} z^* (\sigma)\, a^* (t, \sigma)\, d\sigma.$$

La multiplication à gauche du premier et du second membre de cette relation par e (t) donne

$$e(t)\, \rho^* (t) = \int_{t_0}^{t} e(t)\, z^* (\sigma)\, a^* (t, \sigma)\, d\sigma.$$

En remplaçant le premier et le second membre par leurs espérances mathématiques, on trouve

$$\mathbf{M}\, [e(t)\, \rho^* (t)] = \int_{t_0}^{t} \mathbf{M}\, [e(t)\, z^* (\sigma)]\, a^* (t, \sigma)\, d\sigma.$$

D'après (34), le second membre de la relation obtenue est nul. Par conséquent

$$\mathbf{M}\, [e(t)\, \rho^* (t)] = 0. \tag{27.35}$$

2. Filtres optimaux de Kalman-Bucy. Etudions le système linéaire régi par l'équation différentielle vectorielle

$$\frac{dx}{dt} = A(t)\, x + B(t)\, w(t). \tag{27.36}$$

Ici x est le vecteur de dimension n ; A (t), la matrice $n \times n$; w (t), le processus aléatoire vectoriel de dimension r; B (t), la matrice $n \times r$.

On définit par observation le vecteur z de dimension m décrit par la relation

$$z(t) = C(t)\, x(t) + v(t), \tag{27.37}$$

où C (t) est la matrice $m \times n$; v (t), le processus aléatoire vectoriel de dimension m. Nous avons supposé le système (36), (37) complètement observable avec $w(t) \equiv 0$ et $v(t) \equiv 0$.

Les processus aléatoires w (t) et v (t) sont des bruits blancs gaussiens à moyennes nulles. Leurs matrices de corrélation s'écrivent

$$\begin{aligned} \operatorname{cov} [w(t);\ w(\tau)] &= \mathbf{M}\, [w(t)\, w^* (\tau)] = Q(t)\, \delta(t - \tau), \\ \operatorname{cov} [v(t);\ v(\tau)] &= \mathbf{M}\, [v(t)\, v^* (\tau)] = R(t)\, \delta(t - \tau), \\ \operatorname{cov} [w(t);\ v(\tau)] &= \mathbf{M}\, [w(t)\, v^* (\tau)] = 0, \end{aligned} \tag{27.38}$$

où δ (t) est la fonction impulsion de Dirac ; Q (t), la matrice $r \times r$ symétrique définie non négative ; R (t), la matrice $m \times m$ symétrique définie positive.

Supposons que $x(t_0)$ soit le vecteur aléatoire gaussien de dimension n à moyenne nulle

$$M[x(t_0)] = 0 \tag{27.39}$$

et à matrice de corrélation

$$M[x(t_0)\, x^*(t_0)] = \Sigma_0; \tag{27.40}$$

de plus, la matrice définie non négative Σ_0 est supposée connue. Supposons encore que $w(t)$, $v(t)$ et $x(t_0)$ soient non corrélés.

Remarquons que puisque $w(t)$ est un processus aléatoire à moyenne nulle, (39) entraîne que $M[x(t)] = 0$ pour tout t, c'est-à-dire $x(t)$ est un processus aléatoire à moyenne nulle.

Cherchons le filtre décrit par l'équation différentielle vectorielle

$$\frac{d\rho}{dt} = F(t)\,\rho + G(t)\,z(t), \qquad \rho(t_0) = 0 \tag{27.41}$$

dont la sortie, le vecteur $\rho(t)$ de dimension n, serait une estimation optimale $\hat{x}(t)$ de l'état du système $x(t)$:

$$\hat{x}(t) = \rho(t). \tag{27.42}$$

L'erreur d'estimation

$$e(t) = x(t) - \rho(t) \tag{27.43}$$

peut s'appeler erreur du filtre.

Le signal de sortie du filtre (41) est

$$\rho(t) = \int_{t_0}^{t} \Psi(t, \tau)\, G(\tau)\, z(\tau)\, d\tau, \tag{27.44}$$

où

$$\Psi(t, \tau) = \varkappa(t)\, \varkappa^{-1}(\tau) \tag{27.45}$$

et $\varkappa(t)$ désigne la matrice fondamentale des solutions de l'équation différentielle vectorielle homogène

$$\frac{d\rho}{dt} = F(t)\,\rho. \tag{27.46}$$

Désignons par $a(t, \tau)$ la matrice $n \times m$

$$a(t, \tau) = \Psi(t, \tau)\, G(\tau). \tag{27.47}$$

Constatons que puisque

$$\Psi(t, t) = \varkappa(t)\, \varkappa^{-1}(t) = E,$$

où E est la matrice unité, il vient

$$a(t, t) = G(t). \tag{27.48}$$

Conformément à (47), la sortie du filtre, définie par l'expression (44), peut s'écrire

$$\rho(t)=\int_{t_0}^{t} a(t,\tau)\,z(\tau)\,d\tau. \qquad (27.49)$$

Le choix des paramètres du filtre (41), c'est-à-dire des fonctions $F(t)$ et $G(t)$, doit assurer la minimisation de la fonctionnelle

$$J(t)=\mathbf{M}\,[\langle e(t),\ e(t)\rangle]. \qquad (27.50)$$

Introduisons la notation $\Sigma(t)$ pour la matrice de corrélation de l'erreur du filtre

$$\Sigma(t)=\mathbf{M}\,[e(t)\,e^*(t)]. \qquad (27.51)$$

Notons que puisque d'après (41) $\rho(t_0)=0$, il vient

$$e(t_0)=x(t_0)-\rho(t_0)=x(t_0), \qquad (27.52)$$

et conformément à (40), on a

$$\Sigma(t_0)=\Sigma_0. \qquad (27.53)$$

On voit sans peine que dans notre problème la fonction de poids $a(t,\tau)$ du filtre optimal doit vérifier l'équation intégrale de Wiener (33) obtenue dans ce qui précède:

$$\int_{t_0}^{t} a(t,\tau)\,\mathbf{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau=\mathbf{M}\,[x(t)\,z^*(\sigma)],\quad t_0\leqslant\sigma<t. \qquad (27.54)$$

La fonction de poids matricielle $a(t,\tau)$ qui satisfait à l'équation intégrale matricielle (54) minimise, comme nous l'avons montré au paragraphe précédent, la fonctionnelle (50).

D é t e r m i n a t i o n d e l a f o n c t i o n $F(t)$. En dérivant par rapport à t les deux membres de (54), on obtient

$$\int_{t_0}^{t}\frac{\partial}{\partial t}a(t,\tau)\,\mathbf{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau+a(t,t)\,\mathbf{M}\,[z(t)\,z^*(\sigma)]=$$
$$=\frac{\partial}{\partial t}\mathbf{M}\,[x(t)\,z^*(\sigma)],\qquad t_0\leqslant\sigma<t. \qquad (27.55)$$

Considérons le second membre de cette relation. D'après (36), on a

$$\frac{\partial}{\partial t}\mathbf{M}\,[x(t)\,z^*(\sigma)]=\mathbf{M}\left[\frac{dx(t)}{dt}\,z^*(\sigma)\right]=\mathbf{M}\,\{[A(t)\,x(t)+B(t)\,w(t)]\,z^*(\sigma)\}=$$
$$=A(t)\,\mathbf{M}\,[x(t)\,z^*(\sigma)]+B(t)\,\mathbf{M}\,[w(t)\,z^*(\sigma)]. \qquad (27.56)$$

D'après (37),

$$z^*(\sigma)=x^*(\sigma)\,C^*(\sigma)+v^*(\sigma), \qquad (27.57)$$

donc,

$$\mathbf{M}\,[w(t)\,z^*(\sigma)] = \mathbf{M}\,[w(t)\,x^*(\sigma)]\,C^*(\sigma) + \mathbf{M}\,[w(t)\,v^*(\sigma)],$$

et comme, d'après (38), $\mathbf{M}\,[w(t)\,v^*(\sigma)] = 0$, on obtient

$$\mathbf{M}\,[w(t)\,z^*(\sigma)] = \mathbf{M}\,[w(t)\,x^*(\sigma)]\,C^*(\sigma). \tag{27.58}$$

La solution de l'équation différentielle (36) s'écrit

$$x(\sigma) = \Phi(\sigma, t_0)\,x(t_0) + \int_{t_0}^{\sigma} \Phi(\sigma, \tau)\,B(\tau)\,w(\tau)\,d\tau, \tag{27.59}$$

où

$$\Phi(t, \tau) = \theta(t)\,\theta^{-1}(\tau),$$

et $\theta(t)$ est la matrice fondamentale des solutions de l'équation homogène

$$\frac{dx}{dt} = A(t)\,x.$$

De cette façon,

$$\mathbf{M}\,[w(t)\,x^*(\sigma)] = \mathbf{M}\,[w(t)\,x^*(t_0)]\,\Phi^*(\sigma, t_0) + \\ + \int_{t_0}^{\sigma} \mathbf{M}\,[w(t)\,w^*(\tau)]\,B^*(\tau)\,\Phi^*(\sigma, \tau)\,d\tau. \tag{27.60}$$

Le premier terme du second membre de la relation (60) s'annule en vertu de l'hypothèse suivant laquelle $x(t_0)$ et $w(t)$ sont non corrélés. Le deuxième terme du second membre de (60) contient dans l'expression sous le signe somme le facteur $\mathbf{M}\,[w(t)\,w^*(\tau)] = Q(t)\,\delta(t-\tau)$, où $\tau \in [t_0, \sigma]$. C'est pourquoi pour tout $\sigma \in [t_0, t)$ on a $\tau < t$ et la fonction $\delta(t-\tau)$ s'annule. Le deuxième terme du second membre de (60) est alors nul et la relation (60) devient

$$\mathbf{M}\,[w(t)\,x^*(\sigma)] = 0, \qquad t_0 \leqslant \sigma < t. \tag{27.61}$$

Les relations (61) et (58) entraînent

$$\mathbf{M}\,[w(t)\,z^*(\sigma)] = 0, \qquad t_0 \leqslant \sigma < t. \tag{27.62}$$

Ainsi, conformément à (56) et (62), on a

$$\frac{\partial}{\partial t}\mathbf{M}\,[x(t)\,z^*(\sigma)] = A(t)\,\mathbf{M}\,[x(t)\,z^*(\sigma)], \qquad t_0 \leqslant \sigma < t. \tag{27.63}$$

Examinons maintenant le premier membre de la relation (55). D'après (37)

$$\mathbf{M}\,[z(t)\,z^*(\sigma)] = \mathbf{M}\,\{[C(t)\,x(t) + v(t)]\,z^*(\sigma)\} = \\ = C(t)\,\mathbf{M}\,[x(t)\,z^*(\sigma)] + \mathbf{M}\,[v(t)\,z^*(\sigma)]. \tag{27.64}$$

On a en vertu de (57)

$$\mathbf{M}\,[v(t)\,z^*(\sigma)] = \mathbf{M}\,[v(t)\,x^*(\sigma)]\,C^*(\sigma) + \\ + \mathbf{M}\,[v(t)\,v^*(\sigma)] = \mathbf{M}\,[v(t)\,x^*(\sigma)]\,C^*(\sigma), \qquad t_0 \leqslant \sigma < t, \tag{27.65}$$

du fait que $\mathbf{M}\,[v(t)\,v^*(\sigma)] = 0$ pour $\sigma < t$.

Il s'ensuit de (59) que

$$\mathrm{M}\,[v(t)\,x^*(\sigma)] = \mathrm{M}\,[v(t)\,x^*(t_0)]\,\Phi^*(\sigma, t_0) + \\ + \int_{t_0}^{\sigma} \mathrm{M}\,[v(t)\,w^*(\tau)]\,B^*(\tau)\,\Phi^*(\sigma, \tau)\,d\tau.$$

D'après (38), $\mathrm{M}\,[v(t)\,w^*(\tau)] = 0$, et dans l'hypothèse que $v(t)$ et $x(t_0)$ ne sont pas corrélés, $\mathrm{M}\,[v(t)\,x^*(t_0)] = 0$.

Par conséquent,

$$\mathrm{M}\,[v(t)\,x^*(\sigma)] = 0, \tag{27.66}$$

d'où, d'après (65),

$$\mathrm{M}\,[v(t)\,z^*(\sigma)] = 0, \qquad t_0 \leqslant \sigma < t. \tag{27.67}$$

Ainsi, la relation (64) devient

$$\mathrm{M}\,[z(t)\,z^*(\sigma)] = C(t)\,\mathrm{M}\,[x(t)\,z^*(\sigma)], \qquad t_0 \leqslant \sigma < t. \tag{27.68}$$

Conformément à (63) et à (68), l'équation (55) se met sous la forme

$$[A(t) - a(t, t)\,C(t)]\,\mathrm{M}\,[x(t)\,z^*(\sigma)] - \\ - \int_{t_0}^{t} \frac{\partial}{\partial t} a(t, \tau)\,\mathrm{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau = 0, \qquad t_0 \leqslant \sigma < t. \tag{27.69}$$

La substitution dans l'équation (69) de $\mathrm{M}\,[x(t)\,z^*(\sigma)]$ par sa valeur tirée de (54), donne

$$[A(t) - a(t, t)\,C(t)] \int_{t_0}^{t} a(t, \tau)\,\mathrm{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau - \\ - \int_{t_0}^{t} \frac{\partial}{\partial t} a(t, \tau)\,\mathrm{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau = 0, \qquad t_0 \leqslant \sigma < t,$$

ou

$$\int_{t_0}^{t} [A(t)\,a(t, \tau) - \frac{\partial}{\partial t} a(t, \tau) - a(t, t)\,C(t)\,a(t, \tau)]\,\mathrm{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau = 0, \\ t_0 \leqslant \sigma < t. \tag{27.70}$$

En introduisant les notations

$$b(t, \tau) = A(t)\,a(t, \tau) - \frac{\partial}{\partial t} a(t, \tau) - a(t, t)\,C(t)\,a(t, \tau), \tag{27.71}$$

on ramène l'équation (70) à la forme

$$\int_{t_0}^{t} b(t, \tau)\,\mathrm{M}\,[z(\tau)\,z^*(\sigma)]\,d\tau = 0, \qquad t_0 \leqslant \sigma < t. \tag{27.72}$$

L'addition des premiers et des seconds membres des équations (54) et (72) conduit à l'équation

$$\int_{t_0}^{t} [a(t, \tau)+b(t, \tau)]\, M[z(\tau) z^*(\sigma)]\, d\tau = M[x(t) z^*(\sigma)], \; t_0 \leqslant \sigma < t. \quad (27.73)$$

De la sorte, la fonction $a(t, \tau)$ de même que la fonction $a(t, \tau) + b(t, \tau)$ sont des solutions de l'équation intégrale de Wiener (54), c'est-à-dire ces fonctions minimisent la fonctionnelle (50).

Il est aisé de montrer que puisque la matrice $R(t)$ intervenant dans l'expression (38) est supposée définie positive, la différence entre les deux solutions obtenues de (54) est nulle

$$[a(t, \tau) + b(t, \tau)] - a(t, \tau) = 0, \qquad t_0 \leqslant \tau < t.$$

Pour le démontrer, considérons la relation (34)

$$M[e(t) z^*(\sigma)] = 0, \qquad t_0 \leqslant \sigma < t.$$

Les fonctions $a(t, \tau)$ et $a(t, \tau) + b(t, \tau)$ sont des solutions de (54). D'après (49), à la fonction $a(t, \tau)$ correspond l'estimation optimale

$$\hat{x}(t) = \int_{t_0}^{t} a(t, \tau) z(\tau)\, d\tau.$$

A la fonction $a(t, \tau) + b(t, \tau)$ correspond l'estimation optimale

$$\check{x}(t) = \int_{t_0}^{t} [a(t, \tau)+b(t, \tau)]\, z(\tau)\, d\tau. \quad (27.74)$$

En désignant

$$\varepsilon(t) = \check{x}(t) - \hat{x}(t) \quad (27.75)$$

on aura

$$\varepsilon(t) = \int_{t_0}^{t} b(t, \tau) z(\tau)\, d\tau. \quad (27.76)$$

La condition (34) entraîne les relations

$$M\{[x(t) - \hat{x}(t)]\, z^*(\sigma)\} = 0, \qquad M\{[x(t) - \check{x}(t)]\, z^*(\sigma)\} = 0,$$

d'où l'on déduit que

$$M[\varepsilon(t) z^*(\sigma)] = 0. \quad (27.77)$$

Considérons maintenant les relations (49) et (74) qui conduisent à

$$M[\varepsilon(t) \hat{x}^*(t)] = \int_{t_0}^{t} M[\varepsilon(t) z^*(\sigma)]\, a^*(t, \sigma)\, d\sigma,$$

$$M[\varepsilon(t) \check{x}^*(t)] = \int_{t_0}^{t} M[\varepsilon(t) z^*(\sigma)]\, [a(t, \sigma)+b(t, \sigma)]^*\, d\sigma.$$

Comme d'après (77) les seconds membres de ces relations sont nuls, on a

$$M[\varepsilon(t)\,x^*(t)] = 0, \qquad M[\varepsilon(t)\,\check{x}^*(t)] = 0.$$

Il en résulte que

$$M\{\varepsilon(t)\,[\check{x}(t) - \hat{x}(t)]^*\} = M[\varepsilon(t)\,\varepsilon^*(t)] = 0. \tag{22.78}$$

En portant dans (78) la valeur de $\varepsilon(t)$, on obtient d'après (76)

$$M\left[\int_{t_0}^{t} b(t,\tau)\,z(\tau)\,d\tau \int_{t_0}^{t} z^*(\sigma)\,b^*(t,\sigma)\,d\sigma\right] =$$
$$= \int_{t_0}^{t} b(t,\tau) \int_{t_0}^{t} M[z(\tau)\,z^*(\sigma)]\,b^*(t,\sigma)\,d\sigma\,d\tau = 0. \tag{27.79}$$

Conformément à (37),

$$M[z(\tau)\,z^*(\sigma)] = M\{[y(\tau) + v(\tau)]\,[y^*(\sigma) + v^*(\sigma)]\}, \tag{27.80}$$

où

$$y(t) = C(t)\,x(t). \tag{27.81}$$

Puisque d'après (38), $M[w(t)\,v^*(\tau)] = 0$, les processus aléatoires $x(t)$ et $v(t)$ sont non corrélés. L'expression (80) devient donc

$$M[z(\tau)\,z^*(\sigma)] = M[y(\tau)\,y^*(\sigma)] + M[v(\tau)\,v^*(\sigma)] =$$
$$= M[y(\tau)\,y^*(\sigma)] + R(\tau)\,\delta(\tau - \sigma). \tag{27.82}$$

D'après (82), la relation (79) s'écrit

$$\int_{t_0}^{t} b(t,\tau) \int_{t_0}^{t} M[y(\tau)\,y^*(\sigma)]\,b^*(t,\sigma)\,d\sigma\,d\tau +$$
$$+ \int_{t_0}^{t} b(t,\tau) \int_{t_0}^{t} R(\tau)\,\delta(\tau-\sigma)\,b^*(t,\sigma)\,d\sigma\,d\tau = 0,$$

ou

$$M\left\{\int_{t_0}^{t} b(t,\tau)\,y(\tau)\,d\tau \int_{t_0}^{t} [b(t,\sigma)\,y(\sigma)]^*\,d\sigma\right\} +$$
$$+ \int_{t_0}^{t} b(t,\tau)\,R(\tau)\,b^*(t,\tau)\,d\tau = 0. \tag{27.83}$$

En introduisant la notation

$$\lambda(t) = \int_{t_0}^{t} b(t,\tau)\,y(\tau)\,d\tau, \tag{27.84}$$

réduisons la relation (83) à la forme

$$M[\lambda(t)\,\lambda^*(t)] + \int_{t_0}^{t} b(t,\tau)\,R(\tau)\,b^*(t,\tau)\,d\tau = 0. \tag{27.85}$$

Notons que puisque $x(t)$ est un processus aléatoire à moyenne nulle, les moyennes des processus aléatoires $y(t)$ et $\lambda(t)$ sont nulles elles aussi.

On vérifie facilement que la condition $b(t, \tau) = 0$, $t_0 \leqslant \tau < t$, est suffisante pour que la relation (85) soit réalisée.

Montrons que cette condition est également nécessaire pour que la relation (85) soit réalisée.

Remarquons que les éléments de la matrice $\mathbf{M}[\lambda(t)\lambda^*(t)]$ qui se situent sur sa diagonale principale $\mathbf{M}[\lambda_i^2(t)] \geqslant 0$.

Par hypothèse, la matrice $R(t)$ est symétrique définie positive. Elle peut donc être mise sous la forme

$$R(t) = S(t)S^*(t), \tag{27.86}$$

où $S(t)$ est une matrice réelle régulière. En désignant par $U(t, \tau)$ la matrice $n \times m$

$$U(t, \tau) = b(t, \tau)S(\tau), \tag{27.87}$$

on peut représenter la fonction sous le signe somme de (85) comme suit

$$\Gamma(t, \tau) = b(t, \tau)R(\tau)b^*(t, \tau) = $$

$$= b(t, \tau)S(\tau)S^*(\tau)b^*(t, \tau) = U(t, \tau)U^*(t, \tau). \tag{27.88}$$

Les éléments diagonaux de la matrice $\Gamma(t, \tau)$ de type $n \times n$ sont

$$\Gamma_{ii}(t, \tau) = \sum_{k=1}^{m} U_{ik}^2(t, \tau), \qquad i = 1, \ldots, n. \tag{27.89}$$

Donc, pour annuler les éléments de la diagonale principale de la matrice $n \times n$

$$\mathcal{T}(t, \tau) = \int_{t_0}^{t} b(t, \tau)R(\tau)b^*(t, \tau)\,d\tau, \tag{27.90}$$

il faut remplir la condition

$$U(t, \tau) = b(t, \tau)S(\tau) = 0. \tag{27.91}$$

Puisque

$$b(t, \tau) = U(t, \tau)S^{-1}(\tau), \tag{27.92}$$

la condition (91) devient $b(t, \tau) = 0$.

Si l'on tient compte encore que $\mathbf{M}[\lambda_i^2(t)] \geqslant 0$, où $\lambda(t)$ est définie par l'expression (84), on peut conclure que la condition $b(t, \tau) = 0$, $t_0 \leqslant \tau < t$ est nécessaire pour que les éléments de la diagonale principale de la matrice $n \times n$

$$\mathbf{M}[\lambda(t)\lambda^*(t)] + \int_{t_0}^{t} b(t, \tau)R(\tau)b^*(t, \tau)\,d\tau$$

soient nuls. Pour observer la relation (85), la condition $b(t, \tau) = 0$, $t_0 \leqslant \tau < t$ est donc non seulement suffisante mais encore nécessaire.

Ainsi la relation (85) implique que

$$b(t, \tau) = 0, \qquad t_0 \leqslant \tau < t. \tag{27.93}$$

Il résulte de (93) et (71) que

$$\frac{\partial}{\partial t}a(t, \tau) = A(t)a(t, \tau) - a(t, t)C(t)a(t, \tau). \tag{27.94}$$

Considérons maintenant la relation (49). En dérivant ses deux membres par rapport à t, on obtient

$$\frac{d\rho}{dt}=\int_{t_0}^{t}\frac{\partial}{\partial t}a(t,\tau)z(\tau)\,d\tau+a(t,t)z(t) \qquad (27.95)$$

ou, en vertu de (94),

$$\frac{d\rho}{dt}=\int_{t_0}^{t}[A(t)a(t,\tau)-a(t,t)C(t)a(t,\tau)]\,z(\tau)\,d\tau+a(t,t)z(t)=$$
$$=[A(t)-a(t,t)C(t)]\int_{t_0}^{t}a(t,\tau)z(\tau)\,d\tau+a(t,t)z(t)=$$
$$=[A(t)-a(t,t)C(t)]\,\rho(t)+a(t,t)z(t). \qquad (27.96)$$

Si l'on tient compte que d'après (48), $a(t, t) = G(t)$, et que d'après (41), $\rho(t_0) = 0$, l'équation (96) se met sous la forme

$$\frac{d\rho}{dt}=[A(t)-G(t)C(t)]\,\rho+G(t)z(t), \qquad \rho(t_0)=0. \qquad (27.97)$$

Les équations (97) et (41) montrent que la fonction $F(t)$ de l'équation (41) s'écrit

$$F(t) = A(t) - G(t)C(t). \qquad (27.98)$$

Ainsi le filtre optimal, dont l'équation (97) peut s'écrire

$$\frac{d\rho}{dt}=A(t)\rho+G(t)[z(t)-C(t)\rho], \qquad \rho(t_0)=0, \qquad (27.99)$$

est un système à retour dans lequel l'écart entre avec un gain $G(t)$.

D é t e r m i n a t i o n d e l a f o n c t i o n $G(t)$. Considérons l'équation (54):

$$\mathrm{M}[x(t)z^*(\sigma)]-\int_{t_0}^{t}a(t,\tau)\mathrm{M}[z(\tau)z^*(\sigma)]\,d\tau=0, \quad t_0\leqslant\sigma<t. \qquad (27.100)$$

D'après (37),

$$\mathrm{M}[x(t)z^*(\sigma)] = \mathrm{M}\{x(t)[C(\sigma)x(\sigma)+v(\sigma)]^*\} =$$
$$= \mathrm{M}[x(t)x^*(\sigma)]C^*(\sigma) + \mathrm{M}[x(t)v^*(\sigma)]. \qquad (27.101)$$

Conformément à (59),

$$x(t)=\Phi(t,t_0)x(t_0)+\int_{t_0}^{t}\Phi(t,\tau)B(\tau)w(\tau)\,d\tau, \qquad (27.102)$$

d'où l'on déduit que

$$M[x(t)v^*(\sigma)] = \Phi(t, t_0)M[x(t_0)v^*(\sigma)] + \\ + \int_{t_0}^{t} \Phi(t, \tau)B(\tau)M[w(\tau)v^*(\sigma)]\,d\tau = 0, \quad (27.103)$$

puisque $M[x(t_0)v^*(\sigma)] = 0$ et $M[w(\tau)v^*(\sigma)] = 0$.

Ainsi, l'expression (101) devient

$$M[x(t)z^*(\sigma)] = M[x(t)x^*(\sigma)]C^*(\sigma). \quad (27.104)$$

En vertu de (37)

$$M[z(\tau)z^*(\sigma)] = M\{z(\tau)[C(\sigma)x(\sigma) + v(\sigma)]^*\} = \\ = M[z(\tau)x^*(\sigma)]C^*(\sigma) + M[z(\tau)v^*(\sigma)] = \\ = M[z(\tau)x^*(\sigma)]C^*(\sigma) + \\ + M\{[C(\tau)x(\tau) + v(\tau)]v^*(\sigma)\} = \\ = M[z(\tau)x^*(\sigma)]C^*(\sigma) + R(\tau)\delta(\tau - \sigma), \quad (27.105)$$

comme $M[x(\tau)v^*(\sigma)] = 0$.

En portant (104) et (105) dans (100), il vient

$$M[x(t)x^*(\sigma)]C^*(\sigma) - \int_{t_0}^{t} a(t, \tau)M[z(\tau)x^*(\sigma)]C^*(\sigma)\,d\tau - \\ - \int_{t_0}^{t} a(t, \tau)R(\tau)\delta(\tau - \sigma)\,d\tau = 0, \quad t_0 \leqslant \sigma < t,$$

ou

$$a(t, \sigma)R(\sigma) = M[x(t)x^*(\sigma)]C^*(\sigma) - \\ - \int_{t_0}^{t} a(t, \tau)M[z(\tau)x^*(\sigma)]C^*(\sigma)\,d\tau, \quad t_0 \leqslant \sigma < t. \quad (27.106)$$

Comme les deux membres de la relation (106) sont des fonctions continues de σ, cette relation est vraie également pour $\sigma = t$. On a donc, en tenant compte du fait que $a(t, t) = G(t)$,

$$G(t)R(t) = M[x(t)x^*(t)]C^*(t) - \int_{t_0}^{t} a(t, \tau)M[z(\tau)x^*(t)]C^*(t)\,d\tau,$$

ou, conformément à (49) et (43),

$$G(t)R(t) = M\left\{\left[x(t) - \int_{t_0}^{t} a(t, \tau)z(\tau)\,d\tau\right]x^*(t)\right\}C^*(t) = \\ = M\{[x(t) - \rho(t)]x^*(t)\}C^*(t) = M[e(t)x^*(t)]C^*(t). \quad (27.107)$$

Etant donné que

$$x(t) = e(t) + \rho(t),$$

il vient

$$\mathrm{M}\,[e(t)\,x^*(t)] = \mathrm{M}\{e(t)\,[e(t) + \rho(t)]^*\} = \\ = \mathrm{M}\,[e(t)\,e^*(t)] + \mathrm{M}\,[e(t)\,\rho^*(t)].$$

D'après (35), $\mathrm{M}\,[e(t)\,\rho^*(t)] = 0$. Ainsi, (51) conduit à

$$\mathrm{M}\,[e(t)\,x^*(t)] = \Sigma(t). \qquad (27.108)$$

La relation (107) devient

$$G(t)\,R(t) = \Sigma(t)\,C^*(t). \qquad (27.109)$$

Comme $R(t)$ est une matrice définie positive, on trouve à partir de (109):

$$G(t) = \Sigma(t)\,C^*(t)\,R^{-1}(t). \qquad (27.110)$$

D é t e r m i n a t i o n d e l a f o n c t i o n $\Sigma(t)$. D'après (43),

$$e(t) = x(t) - \rho(t),$$

et donc

$$\frac{de}{dt} = \frac{dx}{dt} - \frac{d\rho}{dt}.$$

Conformément à (36) et (99), cette relation se met sous la forme

$$\frac{de}{dt} = A(t)\,x + B(t)\,w(t) - A(t)\,\rho - G(t)\,[z(t) - C(t)\,\rho] = \\ = A(t)\,e + B(t)\,w(t) - G(t)\,[C(t)\,x + v(t) - C(t)\,\rho],$$

ou

$$\frac{de}{dt} = [A(t) - G(t)\,C(t)]\,e + B(t)\,w(t) - G(t)\,v(t). \qquad (27.111)$$

Etant donné que $w(t)$ et $v(t)$ sont des processus aléatoires à moyennes nulles, il résulte de (111) que $e(t)$ est également un processus aléatoire à moyenne nulle.

D'après (45) et (98), $\varkappa(t)$ est la matrice fondamentale des solutions de l'équation homogène

$$\frac{de}{dt} = [A(t) - G(t)\,C(t)]\,e. \qquad (27.112)$$

Ci-dessus nous avons désigné par $\Psi(t, \tau)$ la matrice (45) de type $n \times n$:

$$\Psi(t, \tau) = \varkappa(t)\,\varkappa^{-1}(\tau).$$

La solution de l'équation (111) peut s'écrire

$$e(t) = \Psi(t, t_0)\,e(t_0) + \int_{t_0}^{t} \Psi(t, \tau)\,[B(\tau)\,w(\tau) - G(\tau)\,v(\tau)]\,d\tau. \qquad (27.113)$$

En vertu de (51) et (113) et compte tenu du fait que $w(t)$ et $v(t)$ sont non corrélés par rapport à $e(t_0) = x(t_0)$, on a

$$\Sigma(t) = \mathrm{M}[e(t)e^*(t)] = \Psi(t, t_0)\mathrm{M}[e(t_0)e^*(t_0)]\Psi^*(t, t_0) + \\ + \mathrm{M}\left\{\int_{t_0}^{t}\Psi(t, \tau)[B(\tau)w(\tau) - G(\tau)v(\tau)]\,d\tau\int_{t_0}^{t}[w^*(\sigma)B^*(\sigma) - \\ - v^*(\sigma)G^*(\sigma)]\Psi^*(t, \sigma)\,d\sigma\right\}. \quad (27.114)$$

Conformément à (53) et (38), l'expression (114) devient

$$\Sigma(t) = \Psi(t, t_0)\Sigma_0\Psi^*(t, t_0) + \int_{t_0}^{t}\Psi(t, \tau)[B(\tau)Q(\tau)B^*(\tau) + \\ + G(\tau)R(\tau)G^*(\tau)]\Psi^*(t, \tau)\,d\tau. \quad (27.115)$$

D'après (110), il faut porter dans l'équation (114) la valeur $G(\tau) = \Sigma(\tau)C^*(\tau)R^{-1}(\tau)$. De la sorte, l'équation (115) présente une équation intégrale matricielle par rapport à la matrice $\Sigma(t)$ de type $n \times n$.

Passons de cette équation à l'équation différentielle équivalente. En dérivant les deux membres de (115) par rapport à t on obtient

$$\frac{d\Sigma}{dt} = \frac{d\Psi(t, t_0)}{dt}\Sigma_0\Psi^*(t, t_0) + \Psi(t, t_0)\Sigma_0\frac{d\Psi^*(t, t_0)}{dt} + \\ + \Psi(t, t)[B(t)Q(t)B^*(t) + G(t)R(t)G^*(t)]\Psi^*(t, t) + \\ + \int_{t_0}^{t}\frac{\partial\Psi(t, \tau)}{\partial t}[B(\tau)Q(\tau)B^*(\tau) + G(\tau)R(\tau)G^*(\tau)]\Psi^*(t, \tau)\,d\tau + \\ + \int_{t_0}^{t}\Psi(t, \tau)[B(\tau)Q(\tau)B^*(\tau) + G(\tau)R(\tau)G^*(\tau)]\frac{\partial\Psi^*(t, \tau)}{\partial t}\,d\tau. \quad (27.116)$$

D'après (98),

$$F(t) = A(t) - G(t)C(t).$$

En vertu de (112) et (98), $\frac{d\varkappa(t)}{dt} = F(t)\varkappa(t)$. C'est pourquoi conformément à (45), on a

$$\frac{\partial\Psi(t, \tau)}{\partial t} = \frac{d\varkappa(t)}{dt}\varkappa^{-1}(\tau) = F(t)\Psi(t, \tau), \\ \frac{\partial\Psi^*(t, \tau)}{\partial t} = \frac{d}{dt}[\varkappa(t)\varkappa^{-1}(\tau)]^* = \Psi^*(t, \tau)F^*(t). \quad (27.117)$$

En retenant que $\Psi(t, t) = E$, où E est une matrice unité, et que d'après (115),

$$\int_{t_0}^{t} \Psi(t, \tau)[B(\tau) Q(\tau) B^*(\tau) + G(\tau) R(\tau) G^*(\tau)] \Psi^*(t, \tau) d\tau =$$
$$= \Sigma(t) - \Psi(t, t_0) \Sigma_0 \Psi^*(t, t_0), \quad (27.118)$$

ramenons l'équation (116) à la forme

$$\frac{d\Sigma}{dt} = F(t) \dot{\Psi}(t, t_0) \Sigma_0 \Psi^*(t, t_0) + \Psi(t, t_0) \Sigma_0 \Psi^*(t, t_0) F^*(t) +$$
$$+ B(t) Q(t) B^*(t) + G(t) R(t) G^*(t) + F(t) [\Sigma(t) - \Psi(t, t_0) \Sigma_0 \Psi^*(t, t_0)] +$$
$$+ [\Sigma(t) - \Psi(t, t_0) \Sigma_0 \Psi^*(t, t_0)] F^*(t),$$

ou

$$\frac{d\Sigma}{dt} = F(t) \Sigma + \Sigma F^*(t) + G(t) R(t) G^*(t) + B(t) Q(t) B^*(t). \quad (27.119)$$

Portons dans l'équation (119) la valeur de $F(t)$ pour obtenir d'après (98)

$$\frac{d\Sigma}{dt} = A(t) \Sigma + \Sigma A^*(t) - G(t) C(t) \Sigma - \Sigma C^*(t) G^*(t) + G(t) R(t) G^*(t) +$$
$$+ B(t) Q(t) B^*(t). \quad (27.120)$$

Si l'on substitue à $G(t)$ sa valeur tirée de (110) et si l'on tient compte que d'après (53) $\Sigma(t_0) = \Sigma_0$, on obtient l'équation différentielle

$$\frac{d\Sigma}{dt} = A(t) \Sigma + \Sigma A^*(t) - \Sigma C^*(t) R^{-1}(t) C(t) \Sigma +$$
$$+ B(t) Q(t) B^*(t), \quad \Sigma(t_0) = \Sigma_0, \quad (27.121)$$

que vérifie la matrice $\Sigma(t)$ de type $n \times n$.

L'équation (121) est une équation différentielle matricielle de Riccati.

Maintenant d'après (110) on peut écrire l'équation différentielle (99) du filtre optimal sous la forme

$$\frac{d\rho}{dt} = [A(t) - \Sigma(t) C^*(t) R^{-1}(t) C(t)] \rho +$$
$$+ \Sigma(t) C^*(t) R^{-1}(t) z(t), \quad \rho(t_0) = 0, \quad (27.122)$$

où $\Sigma(t)$ est une solution de l'équation (121).

3. Méthode de résolution par approximations successives de l'équation différentielle matricielle de Riccati. L'équation de Riccati (121) s'écrit

$$\frac{d\Sigma}{dt} = A(t) \Sigma + \Sigma A^*(t) - \Sigma C^*(t) R^{-1}(t) C(t) \Sigma +$$
$$+ B(t) Q(t) B^*(t). \qquad \Sigma(t_0) = \Gamma \geqslant 0.$$

Ici $R(t)$ est la matrice symétrique définie positive, $Q(t)$ et Γ des matrices symétriques définies non négatives.

Introduisons les notations

$$\begin{aligned} \mathfrak{R}(t) &= C^*(t)\, R^{-1}(t)\, C(t), \\ L(t) &= B(t)\, Q(t)\, B^*(t). \end{aligned} \qquad (27.123)$$

L'équation (121) devient

$$\frac{d\Sigma}{dt} = A(t)\,\Sigma + \Sigma A^*(t) - \Sigma\mathfrak{R}(t)\,\Sigma + L(t), \quad \Sigma(t_0) = \Gamma \geqslant 0. \qquad (27.124)$$

Pour résoudre l'équation (124), on peut appliquer le processus itératif suivant. Les approximations successives de la solution de (124) $\Sigma^{(0)}(t)$, $\Sigma^{(1)}(t)$, $\Sigma^{(2)}(t)$, ..., $\Sigma^{(n-1)}(t)$, $\Sigma^{(n)}(t)$, ... s'obtiennent à partir des équations différentielles *linéaires*

$$\frac{d\Sigma^{(0)}}{dt} = A(t)\,\Sigma^{(0)} + \Sigma^{(0)} A^*(t) + L(t), \quad \Sigma^{(0)}(t_0) = \Gamma \geqslant 0,$$

$$\frac{d\Sigma^{(1)}}{dt} = [A(t) - \Sigma^{(0)}\mathfrak{R}(t)]\,\Sigma^{(1)} + \Sigma^{(1)}[A^*(t) - \mathfrak{R}(t)\,\Sigma^{(0)}] + \Sigma^{(0)}\mathfrak{R}(t)\,\Sigma^{(0)} + L(t), \quad \Sigma^{(1)}(t_0) = \Gamma \geqslant 0,$$

$$\frac{d\Sigma^{(2)}}{dt} = [A(t) - \Sigma^{(1)}\mathfrak{R}(t)]\,\Sigma^{(2)} + \Sigma^{(2)}[A^*(t) - \mathfrak{R}(t)\,\Sigma^{(1)}] + \Sigma^{(1)}\mathfrak{R}(t)\,\Sigma^{(1)} + L(t), \quad \Sigma^{(2)}(t_0) = \Gamma \geqslant 0,$$

. . . .

$$\frac{d\Sigma^{n-1}}{dt} = [A(t) - \Sigma^{(n-2)}\mathfrak{R}(t)]\,\Sigma^{(n-1)} + \Sigma^{(n-1)}[A^*(t) - \mathfrak{R}(t)\,\Sigma^{(n-2)}] + \Sigma^{(n-2)}\mathfrak{R}(t)\,\Sigma^{(n-2)} + L(t), \quad \Sigma^{(n-1)}(t_0) = \Gamma \geqslant 0,$$

$$\frac{d\Sigma^{(n)}}{dt} = [A(t) - \Sigma^{(n-1)}\mathfrak{R}(t)]\,\Sigma^{(n)} + \Sigma^{(n)}[A^*(t) - \mathfrak{R}(t)\,\Sigma^{(n-1)}] + \Sigma^{(n-1)}\mathfrak{R}(t)\,\Sigma^{(n-1)} + L(t), \quad \Sigma^{(n)}(t_0) = \Gamma \geqslant 0,$$

. . . .

(27.125)

Introduisons les notations

$$\Delta_n = \Sigma^{(n)} - \Sigma^{(n-1)}, \qquad (27.126)$$

$$\widetilde{A}_n = A(t) - \Sigma^{(n-1)}\mathfrak{R}(t). \qquad (27.127)$$

Constatons que puisque $\Sigma^{(n)}$ et $\Sigma^{(n-1)}$ sont des matrices symétriques, Δ_n est également une matrice symétrique.

Avec $n \geqslant 2$, des équations différentielles que vérifient $\Sigma^{(n)}$ et $\Sigma^{(n-1)}$ on trouve que Δ_n satisfait à l'équation différentielle suivante

$$\frac{d\Delta_n}{dt} = \widetilde{A}_n \Delta_n + \Delta_n \widetilde{A}_n^* - \Delta_{n-1}\mathfrak{R}(t)\,\Delta_{n-1}, \quad n \geqslant 2, \quad \Delta_n(t_0) = 0. \qquad (27.128)$$

Remarquons encore que conformément à (126) l'équation différentielle (125) vérifiée par la matrice $\Sigma^{(n)}$ peut se mettre sous la forme

$$\frac{d\Sigma^{(n)}}{dt} = A(t)\,\Sigma^{(n)} + \Sigma^{(n)} A^*(t) + \Delta_n(t)\,\mathfrak{R}(t)\,\Delta_n(t) - \Sigma^{(n)}\mathfrak{R}(t)\,\Sigma^{(n)} + L(t), \quad \Sigma^{(n)}(t_0) = \Gamma \geqslant 0. \qquad (27.129)$$

L'équation différentielle matricielle (129) est équivalente à l'équation intégrale matricielle

$$\Sigma^{(n)}(t) = \Phi(t, t_0)\,\Gamma\Phi^*(t, t_0) + \int_{t_0}^{t} \Phi(t, \tau)\,[L(\tau) - \Sigma^{(n)}(\tau)\,\mathfrak{R}(\tau)\,\Sigma^{(n)}(\tau) + \Delta_n(\tau)\,\mathfrak{R}(\tau)\,\Delta_n(\tau)]\,\Phi^*(t, \tau)\,d\tau, \qquad (27.130)$$

où

$$\Phi(t, \tau) = \theta(t)\,\theta^{-1}(\tau),$$

et $\theta(t)$ est la matrice fondamentale des solutions de l'équation différentielle vectorielle

$$\frac{dx}{dt} = A(t)\,x.$$

D'une façon analogue, l'équation différentielle vectorielle de Riccati (124) est équivalente à l'équation intégrale matricielle

$$\Sigma(t) = \Phi(t, t_0)\,\Gamma\Phi^*(t, t_0) + \int_{t_0}^{t} \Phi(t, \tau)\,[L(\tau) - \Sigma(\tau)\,\mathfrak{R}(\tau)\,\Sigma(\tau)]\,\Phi^*(t, \tau)\,d\tau. \qquad (27.131)$$

En introduisant les notations

$$\Psi_n(t, \tau) = \varkappa_n(t)\,\varkappa_n^{-1}(\tau),$$

où $\varkappa_n(t)$ est la matrice fondamentale des solutions de l'équation différentielle vectorielle

$$\frac{d\rho_n}{dt} = \widetilde{A}_n(t)\,\rho_n$$

et en tenant compte que $\Delta_n(t_0) = 0$, cherchons la solution de l'équation différentielle matricielle (128) sous la forme

$$\Delta_n(t) = -\int_{t_0}^{t} \Psi_n(t, \tau)\,\Delta_{n-1}(\tau)\,\mathfrak{R}(\tau)\,\Delta_{n-1}(\tau)\,\Psi_n^*(t, \tau)\,d\tau \leqslant 0, \quad n \geqslant 2. \qquad (27.132)$$

Dans ces conditions, pour tout t et $n \geqslant 1$, $\Sigma^{(n)}(t) \geqslant 0$, du fait que d'après (125)

$$\Sigma^{(n)}(t) = \Psi_n(t, t_0)\,\Gamma\Psi_n^*(t, t_0) + \int_{t_0}^{t} \Psi_n(t, \tau)\,[L(\tau) + \Sigma^{(n-1)}(\tau)\,\mathfrak{R}(\tau)\,\Sigma^{(n-1)}(\tau)]\,\Psi_n^*(t, \tau)\,d\tau. \qquad (27.133)$$

Théorème. *Soit la matrice* $\Sigma^{(0)}(t) \geqslant 0$ *pour tout* t *et dérivable par rapport à* t. *Dans ce cas, sur chaque intervalle fini* (t_0, T) *on a la relation* $\Sigma^{(n)}(t) \leqslant \Sigma^{(n-1)}(t)$, $n > 1$, *et pour* $n \to \infty$ *la suite des matrices* $\{\Sigma^{(n)}(t)\}$ *converge uniformément vers la solution* $\Sigma(t)$ *de l'équation de Riccati* (124).

Démonstration. Il s'ensuit de (125) que

$$\Sigma^{(0)}(t) = \Phi(t, t_0)\,\Gamma\Phi^*(t, t_0) + \int_{t_0}^{t} \Phi(t, \tau)\,L(\tau)\,\Phi^*(t, \tau)\,d\tau, \qquad (27.134)$$

et puisque $L(t)$ est une matrice symétrique définie non négative on a $\Sigma^{(0)}(t) \geqslant 0$ pour tout $t \in (t_0, T)$, comme le prévoient les conditions du théorème.

D'après (132), $\Delta_n(t) \leqslant 0$ pour $n \geqslant 2$, soit $\Sigma^{(n)}(t) \leqslant \Sigma^{(n-1)}(t)$ pour $n \geqslant 2$ et pour tout t. Ainsi, on a la suite des matrices

$$0 \leqslant \ldots \leqslant \Sigma^{(n)}(t) \leqslant \Sigma^{(n-1)}(t) \leqslant \ldots \leqslant \Sigma^{(2)}(t) \leqslant \Sigma^{(1)}(t).$$

(Constatons que l'inégalité $\Sigma^{(0)}(t) \geqslant \Sigma^{(1)}(t)$ peut ne pas avoir lieu.)

Comme d'après (125)

$$\Sigma^{(1)}(t) = \Psi_1(t, t_0)\,\Gamma\Psi_1^*(t, t_0) + \int_{t_0}^{t} \Psi_1(t, \tau)\,[\Sigma^{(0)}(\tau)\,\mathfrak{R}(\tau)\,\Sigma^{(0)}(\tau) + L(\tau)]\,\Psi_1^*(t, \tau)\,d\tau, \qquad (27.135)$$

on a $\sup\limits_{t\in[t_0,\,T]} \|\Sigma^{(1)}(t)\| \leqslant K < \infty$. On en tire suivant le théorème de la convergence monotone des opérateurs positifs dans un espace de Hilbert [38] que $\lim\limits_{n\to\infty} \Sigma^{(n)}(t) = \overline{\Sigma}(t)$, $t \in [t_0, T]$, où la matrice $\overline{\Sigma}(t)$ vérifie l'équation intégrale

$$\overline{\Sigma}(t) = \Phi(t, t_0)\,\Gamma\Phi^*(t, t_0) + \int_{t_0}^{t} \Phi(t, \tau)\,[L(\tau) - \overline{\Sigma}(\tau)\,\mathfrak{R}(\tau)\,\overline{\Sigma}(\tau)]\,\Phi^*(t, \tau)\,d\tau \qquad (27.136)$$

qui diffère de l'équation intégrale (130) vérifiée par la matrice $\Sigma^{(n)}(t)$ par l'absence dans le second membre du terme

$$\int_{t_0}^{t} \Phi(t, \tau)\,\Delta_n(\tau)\,\mathfrak{R}(\tau)\,\Delta_n(\tau)\,\Phi^*(t, \tau)\,d\tau, \qquad (27.137)$$

où $\Delta_n(t)\,\mathfrak{R}(t)\,\Delta_n(t)$ est une matrice symétrique définie non négative.

(136) et (131) impliquent que

$$\overline{\Sigma}(t) = \Sigma(t), \qquad (27.138)$$

où $\Sigma(t)$ est une solution d e l'équation de Riccati (124). Le théorème est démontré

§ 28. Application du principe du maximum de Pontriaguine à la théorie des filtres optimaux de Kalman-Bucy

Au § 27 nous avons construit les filtres optimaux de Kalman-Bucy sur la base de l'équation intégrale de Wiener. Dans ce qui suit, sous des hypothèses quelques peu plus générales sur la fonctionnelle qui donne la performance du filtrage, nous exposons le mode de construction des filtres de Kalman-Bucy basé sur le principe du maximum de Pontriaguine.

Considérons le système linéaire non stationnaire

$$\frac{dx}{dt} = A(t)\,x + B(t)\,w(t), \qquad y = C(t)\,x. \qquad (28.1)$$

Ici x est le vecteur de dimension n d'état du système ; $w(t)$, le vecteur de dimension r qui est une entrée du système; $y(t)$, le vecteur de dimension m qui détermine la sortie du système ; $A(t)$, la matrice $n \times n$; $B(t)$, la matrice $n \times r$; $C(t)$, la matrice $m \times n$. Le système (I) est supposé *complètement observable.*

L'entrée $w(t)$ est un processus aléatoire vectoriel gaussien du type bruit blanc à moyenne nulle

$$\mathrm{M}\,[w(t)] = 0 \quad \text{pour tout } t \tag{28.2}$$

et à matrice de corrélation

$$\operatorname{cov}\,[w(t);\ w(\tau)] = \mathrm{M}\,[w(t)\,w^*(\tau)] = \delta(t-\tau)\,Q(t), \tag{28.3}$$

où $\delta(t)$ est la fonction impulsion de Dirac, et $Q(t)$, la matrice symétrique définie non négative $r \times r$.

Soit t_0 l'instant initial ($t_0 > -\infty$), et le vecteur $x(t_0)$ l'état initial du système. On suppose que $x(t_0)$ est une variable aléatoire vectorielle gaussienne indépendante de $w(t)$ à moyenne connue

$$\mathrm{M}\,[x(t_0)] = \bar{x}_0, \tag{28.4}$$

et à matrice de corrélation connue

$$\operatorname{cov}\,[x(t_0);\ x(t_0)] = \mathrm{M}\,\{[x(t_0)-\bar{x}_0]\,[x(t_0)-\bar{x}_0]^*\} = \Sigma_0. \tag{28.5}$$

Sous les hypothèses que nous venons de faire ci-dessus, $x(t)$ et $y(t)$ sont des processus aléatoires gaussiens.

Supposons que l'observation du vecteur $y(t)$ ait lieu en présence des perturbations du type bruit blanc gaussien et, de ce fait, le signal observé est de la forme

$$z(t) = y(t) + v(t) = C(t)\,x(t) + v(t), \tag{28.6}$$

où $v(t)$ est un bruit blanc gaussien à moyenne nulle

$$\mathrm{M}\,[v(t)] = 0 \quad \text{pour tout } t \tag{28.7}$$

et à matrice de corrélation

$$\operatorname{cov}\,[v(t);\ v(\tau)] = \mathrm{M}\,[v(t)\,v^*(\tau)] = \delta(t-\tau)\,R(t), \tag{28.8}$$

avec $R(t)$, la matrice symétrique définie positive $m \times m$. Supposons de plus que $v(t)$, $w(t)$ et $x(t_0)$ soient non corrélés et, par conséquent, indépendants.

On demande d'obtenir l'estimation $\hat{x}(t)$ de l'état $x(t)$ du système (I) à partir du vecteur fonction $z(t)$ observable dans l'intervalle de temps $[t_0, T]$. (Les vecteurs fonctions $x(t)$ et $y(t)$ ne sont pas mesurables directement.) Appelons *filtre* le système qui détermine le vecteur $\hat{x}(t)$. Nous allons nous borner ici aux filtres qui constituent des systèmes linéaires non stationnaires de la forme

$$\frac{d\rho}{dt} = F(t)\,\rho + G(t)\,z(t), \tag{28.9}$$

où $\rho(t)$ est le vecteur de dimension n; $F(t)$, la matrice $n \times n$; $G(t)$, la matrice $n \times m$. Le vecteur fonction $\rho(t)$ est adopté comme estimation $\hat{x}(t)$ de l'état $x(t)$ du système initial (I)

$$\hat{x}(t) = \rho(t). \qquad (28.10)$$

Conformément à (10), l'erreur d'estimation $e(t) = x(t) - \hat{x}(t)$ s'écrit

$$e(t) = x(t) - \rho(t). \qquad (28.11)$$

Le vecteur fonction $e(t)$ peut être appelé erreur du filtre.

Voici les contraintes auxquelles doit satisfaire le filtre (9): le vecteur fonction $\rho(t)$ doit être une estimation *non déplacée* de $x(t)$, et de plus, minimiser la variance de l'erreur (plus précisément, minimiser une certaine fonctionnelle qui dépend de la matrice de corrélation du processus aléatoire vectoriel $e(t)$). Le choix des matrices $F(t)$ et $G(t)$ ainsi que celui de l'état initial du filtre $\rho(t_0)$ doivent être guidés par ces contraintes.

Cherchons d'abord l'équation différentielle vérifiée par le vecteur fonction $e(t)$. D'après (11), (1), (9) et (6), on a

$$\frac{de}{dt} = [A(t) - F(t) - G(t)C(t)]\,x(t) + \\ + F(t)\,e(t) + B(t)\,w(t) - G(t)\,v(t). \qquad (28.12)$$

Afin que le vecteur fonction $\rho(t)$ soit une estimation non déplacée de $x(t)$, il faut observer la condition

$$\mathrm{M}[x(t)] = \mathrm{M}[\rho(t)] \text{ pour tout } t \geqslant t_0, \qquad (28.13)$$

d'où l'on déduit d'après (11)

$$\mathrm{M}[e(t)] = 0 \text{ pour tout } t \geqslant t_0, \qquad (28.14)$$

$$\frac{d}{dt}\mathrm{M}[e(t)] = \mathrm{M}\left[\frac{de}{dt}\right] = 0 \text{ pour tout } t \geqslant t_0. \qquad (28.15)$$

En tenant compte du fait que d'après (2) et (7), $\mathrm{M}[w(t)] = 0$, $\mathrm{M}[v(t)] = 0$, et en remplaçant le premier et le second membre de l'équation (12) par leurs espérances mathématiques, on obtient en vertu de (14) et (15) la relation

$$[A(t) - F(t) - G(t)C(t)]\,\mathrm{M}[x(t)] = 0. \qquad (28.16)$$

Comme $\mathrm{M}[x(t)] \neq 0$, la condition (16) de l'estimation non déplacée de $x(t)$ devient

$$A(t) - F(t) - G(t)C(t) = 0. \qquad (28.17)$$

De cette façon, les matrices $F(t)$ et $G(t)$, qui font partie de l'équation du filtre (9), doivent vérifier la relation

$$F(t) = A(t) - G(t)C(t). \qquad (28.18)$$

Conformément à (13) et (4), l'état initial du filtre $\rho(t_0)$ doit être

$$\rho(t_0) = \overline{x_0}. \qquad (28.19)$$

Si les conditions (18) et (19) sont remplies, le vecteur fonction $\rho(t)$ est pour tout $t \geqslant t_0$ une estimation non déplacée de l'état $x(t)$ du système initial (1).

Ainsi le filtre fournissant l'estimation non déplacée de $x(t)$ est décrit par l'équation différentielle

$$\frac{d\rho}{dt} = [A(t) - G(t)C(t)]\rho + G(t)z(t), \quad \rho(t_0) = \bar{x}_0. \tag{28.20}$$

Les filtres décrits par l'équation (20) comportent comme paramètre la matrice $G(t)$. La matrice $G(t)$ doit être choisie en se guidant par la condition de minimisation de la variance de l'erreur $e(t)$.

Indiquons maintenant le critère de performance du filtre. Remarquons au préalable que puisque d'après (14) $\mathbf{M}[e(t)] = 0$, la matrice de corrélation de l'erreur $e(t)$ est

$$\Sigma(t) = \mathbf{M}[e(t)e^*(t)]. \tag{28.21}$$

La matrice $\Sigma(t)$ est symétrique. Sa valeur initiale $\Sigma(t_0)$ s'écrit

$$\Sigma(t_0) = \mathbf{M}[e(t_0)e^*(t_0)] = \\ = \mathbf{M}\{[x(t_0) - \rho(t_0)][x(t_0) - \rho(t_0)]^*\} = \\ = \mathbf{M}\{[x(t_0) - \bar{x}_0][x(t_0) - \bar{x}_0]^*\},$$

ou, d'après (5),

$$\Sigma(t_0) = \Sigma_0. \tag{28.22}$$

Comme nous l'avons dit plus haut, la matrice Σ_0 est supposée connue.

Considérons la fonctionnelle

$$J = \mathbf{M}[e^*(T)L(T)e(T)], \tag{28.23}$$

où T est l'instant fixé, et $L(t)$ la matrice symétrique définie positive $n \times n$. Puisque le produit scalaire a^*b de deux vecteurs a et b de dimension n est égal à la trace de la matrice ba^* de type $n \times n$, c'est-à-dire à la somme des éléments diagonaux,

$$a^*b = \mathrm{Sp}[ba^*],$$

l'expression (23) peut s'écrire

$$J = \mathbf{M}\{\mathrm{Sp}[L(T)e(T)e^*(T)]\} = \\ = \mathrm{Sp}\{L(T)\mathbf{M}[e(T)e^*(T)]\},$$

ou, en vertu de (21),

$$J = \mathrm{Sp}[L(T)\Sigma(T)]. \tag{28.24}$$

Adoptons comme mesure de la performance du filtre (20) la fonctionnelle (24). Dans l'équation différentielle (20) la matrice $G(t)$ doit être choisie de façon à minimiser la fonctionnelle (24). Le filtre qui satisfait à cette condition s'appelle *filtre optimal*.

Cherchons maintenant l'équation différentielle vérifiée par la matrice de corrélation $\Sigma(t)$.

L'équation différentielle (12) vérifiée par le vecteur fonction $e(t)$ devient, en vertu de (17) et (18),

$$\frac{de}{dt}=[A(t)-G(t)C(t)]e(t)+B(t)w(t)-G(t)v(t). \qquad (28.25)$$

Puisque d'après (21)

$$\Sigma(t)=\mathbf{M}[e(t)e^*(t)],$$

il vient

$$\frac{d\Sigma}{dt}=\frac{d}{dt}\mathbf{M}[e(t)e^*(t)]=\mathbf{M}\left[\frac{de}{dt}e^*(t)\right]+\mathbf{M}\left[e(t)\frac{de^*}{dt}\right],$$

ou, d'après (25),

$$\frac{d\Sigma}{dt}=[A(t)-G(t)C(t)]\mathbf{M}[e(t)e^*(t)]+\mathbf{M}[e(t)e^*(t)][A(t)-G(t)C(t)]^*+$$
$$+B(t)\mathbf{M}[w(t)e^*(t)]+\mathbf{M}[e(t)w^*(t)]B^*(t)-$$
$$-G(t)\mathbf{M}[v(t)e^*(t)]-\mathbf{M}[e(t)v^*(t)]G^*(t). \qquad (28.26)$$

En tenant compte de (21), l'équation (26) peut s'écrire

$$\frac{d\Sigma}{dt}=[A(t)-G(t)C(t)]\Sigma(t)+\Sigma(t)[A(t)-G(t)C(t)]^*+$$
$$+B(t)\mathbf{M}[w(t)e^*(t)]+\mathbf{M}[e(t)w^*(t)]B^*(t)-$$
$$-G(t)\mathbf{M}[v(t)e^*(t)]-\mathbf{M}[e(t)v^*(t)]G^*(t). \qquad (28.27)$$

Remarquons maintenant que d'après (2), (7), (14),

$$\mathbf{M}[w(t)]=0, \qquad \mathbf{M}[v(t)]=0, \qquad \mathbf{M}[e(t)]=0.$$

Conformément à (11) et à (19),

$$e(t_0)=x(t_0)-\bar{x}_0. \qquad (28.28)$$

Puisqu'on suppose que $v(t)$, $w(t)$ et $x(t_0)$ sont indépendants (non corrélés), on en tire que

$$\begin{aligned}&\mathbf{M}[e(t_0)w^*(t)]=0, \qquad &&\mathbf{M}[e(t_0)v^*(t)]=0,\\ &\mathbf{M}[w(t)e^*(t_0)]=0, \qquad &&\mathbf{M}[v(t)e^*(t_0)]=0.\end{aligned} \qquad (28.29)$$

Désignons maintenant par $\Psi(t, \tau)$ la matrice $n \times n$

$$\Psi(t, \tau)=\varkappa(t)\varkappa^{-1}(\tau), \qquad (28.30)$$

où $\varkappa(t)$ est la matrice fondamentale des solutions de l'équation différentielle homogène

$$\frac{de}{dt}=[A(t)-G(t)C(t)]e(t). \qquad (28.31)$$

La solution de l'équation différentielle (25) est de la forme

$$e(t)=\Psi(t, t_0)e(t_0)+\int_{t_0}^{t}\Psi(t, \tau)B(\tau)w(\tau)d\tau-$$
$$-\int_{t_0}^{t}\Psi(t, \tau)G(\tau)v(\tau)d\tau. \qquad (28.32)$$

En vertu de (32), (29), (3) et (8), on a

$$\mathbf{M}\,[e(t)\,w^*(t)] = \Psi(t,\ t_0)\,\mathbf{M}\,[e(t_0)\,w^*(t)] + \int_{t_0}^{t} \Psi(t,\ \tau)\,B(\tau)\,\mathbf{M}\,[w(\tau)\,w^*(t)]\,d\tau - \int_{t_0}^{t} \Psi(t,\ \tau)\,G(\tau)\,\mathbf{M}\,[v(\tau)\,w^*(t)]\,d\tau = \int_{t_0}^{t} \Psi(t,\ \tau)\,B(\tau)\,\delta(t-\tau)\,Q(t)\,d\tau = \frac{1}{2}\,B(t)\,Q(t). \tag{28.33}$$

$$-\mathbf{M}\,[e(t)\,v^*(t)] = -\Psi(t,\ t_0)\,\mathbf{M}\,[e(t_0)\,v^*(t)] - \int_{t_0}^{t} \Psi(t,\ \tau)\,B(\tau)\,\mathbf{M}\,[w(\tau)\,v^*(t)]\,d\tau + \int_{t_0}^{t} \Psi(t,\ \tau)\,G(\tau)\,\mathbf{M}\,[v(\tau)\,v^*(t)]\,d\tau = \int_{t_0}^{t} \Psi(t,\ \tau)\,G(\tau)\,\delta(t-\tau)\,R(\tau)\,d\tau = \frac{1}{2}\,G(t)\,R(t). \tag{28.34}$$

D'une façon analogue

$$\mathbf{M}\,[w(t)\,e^*(t)] = \frac{1}{2}\,Q(t)\,B^*(t), \tag{28.35}$$

$$-\mathbf{M}\,[v(t)\,e^*(t)] = \frac{1}{2}\,R(t)\,G^*(t). \tag{28.36}$$

Les expressions (27), (33)-(36) et (22) font que la matrice $\Sigma\ (t)$ est définie par l'équation différentielle et la condition initiale

$$\frac{d\Sigma}{dt} = [A(t) - G(t)\,C(t)]\,\Sigma + \Sigma\,[A(t) - G(t)\,C(t)]^* + B(t)\,Q(t)\,B^*(t) + G(t)\,R(t)\,G^*(t),\quad \Sigma(t_0) = \Sigma_0. \tag{28.37}$$

Comme les matrices $A\ (t)$, $B\ (t)$, $C\ (t)$, $Q\ (t)$, $R\ (t)$ et Σ_0 sont données, le paramètre dont dépend la solution $\Sigma\ (t)$ du problème de Cauchy (37) est la matrice $G\ (t)$. Ainsi, la valeur de la fonctionnelle (24)

$$J = \mathrm{Sp}\,[L\,(T)\ \Sigma\,(T)]$$

dépend également de la matrice $G\,(t)$ comme d'un paramètre.

La matrice $\hat{G}\,(t)$ qui minimise la fonctionnelle (24) est *optimale*, et le filtre (20) dont $G\ (t) = \hat{G}\ (t)$ est un *filtre optimal*. Dans l'ouvrage de base de Kalman et Bucy [37], le filtre optimal se construit à l'aide des méthodes de calcul fonctionnel. Dans ce qui suit nous donnons une autre méthode de résolution du problème [5].

Si l'on considère l'ensemble des éléments σ_{ij} $(i, j = 1, \ldots, n)$ de la matrice Σ comme un *état* du système (37) et les éléments $\tilde{g}_{jk}$ $(j = 1, \ldots, n;\ k = 1, \ldots, m)$ de la matrice G comme des *commandes* appliquées à ce système, on aboutit, comme l'a montré M. Athans [5], à l'un des problèmes déterministes de commande optimale qui consiste à choisir les commandes $\tilde{g}_{jk}(t)$ $(j = 1, \ldots, n;\ k = 1, \ldots, m)$ $(t_0 \leqslant t \leqslant T)$ de façon à minimiser la fonctionnelle (24); de plus, dans ce problème le temps T est fixé et l'extrémité de la trajectoire est libre. Ce problème peut être résolu en appliquant le principe du maximum de Pontriaguine.

D'une façon analogue à (24.5) la commande optimale $\hat{G}(t)$ doit minimiser la fonction

$$\mathscr{H} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{d\sigma_{ij}}{dt} p_{ij}(t) = \operatorname{Sp}\left[\frac{d\Sigma}{dt} P^*(t)\right], \tag{28.38}$$

où par $P(t)$ on désigne la matrice $n \times n$ dont les éléments p_{ij} $(i, j = 1, \ldots, n)$ sont de même que dans (24.4) des variables auxiliaires du problème à temps fixé et à extrémité libre de la trajectoire.

Conformément à (37), l'expression (38) peut être ramenée à la forme

$$\begin{aligned}\mathscr{H} = {} & \operatorname{Sp}[A(t)\Sigma(t)P^*(t)] - \operatorname{Sp}[G(t)C(t)\Sigma(t)P^*(t)] + \\ & + \operatorname{Sp}[\Sigma(t)A^*(t)P^*(t)] - \operatorname{Sp}[\Sigma(t)C^*(t)G^*(t)P^*(t)] + \\ & + \operatorname{Sp}[B(t)Q(t)B^*(t)P^*(t)] + \operatorname{Sp}[G(t)R(t)G^*(t)P^*(t)]. \end{aligned} \tag{28.39}$$

Proposons-nous maintenant de déduire certaines relations auxiliaires dont nous aurons besoin dans ce qui suit.

1) Considérons la matrice $axbx^*$, où a est la matrice $n \times n$; b, la matrice $m \times m$; x, la matrice $n \times m$. Cherchons la trace de cette matrice $\operatorname{Sp}[axbx^*]$, ainsi que les dérivées partielles de $\operatorname{Sp}[axbx^*]$ par rapport à x_{jk} $(j = 1, \ldots, n;\ k = 1, \ldots, m)$, où x_{jk} sont les éléments de la matrice x.

Introduisons les notations

$$c = ax,\quad d = bx^*,\quad e = cd.$$

Les éléments des matrices c et d sont

$$c_{\rho\nu} = \sum_{l=1}^{n} a_{\rho l} x_{l\nu},\qquad d_{\nu\rho} = \sum_{\mu=1}^{m} b_{\nu\mu} x_{\rho\mu} \qquad (\rho = 1, \ldots, n;\ \nu = 1, \ldots, m).$$

D'où

$$e_{\rho\rho} = \sum_{\nu=1}^{m} c_{\rho\nu} d_{\nu\rho}$$

ou

$$e_{\rho\rho} = \sum_{\nu=1}^{m} \left(\sum_{l=1}^{n} a_{\rho l} x_{l\nu}\right)\left(\sum_{\mu=1}^{m} b_{\nu\mu} x_{\rho\mu}\right) \qquad (\rho = 1, \ldots, n).$$

La trace de la matrice $e = axbx^*$ s'écrit

$$\mathrm{Sp}\,[axbx^*] = \sum_{\rho=1}^{n} \sum_{\nu=1}^{m} \Big(\sum_{l=1}^{n} a_{\rho l} x_{l\nu}\Big) \Big(\sum_{\mu=1}^{m} b_{\nu\mu} x_{\rho\mu}\Big). \tag{28.40}$$

L'expression (40) conduit à

$$\frac{\partial}{\partial x_{jk}} \mathrm{Sp}\,[axbx^*] = \sum_{\rho=1}^{n} \sum_{\mu=1}^{m} a_{\rho j} x_{\rho\mu} b_{k\mu} + \sum_{l=1}^{n} \sum_{\nu=1}^{m} a_{jl} x_{l\nu} b_{\nu k}$$

$$(j = 1, \ldots, n\,;\ k = 1, \ldots, m). \tag{28.41}$$

La matrice rectangulaire $n \times m$ dont les éléments sont

$$\frac{\partial}{\partial x_{jk}} \mathrm{Sp}\,[axbx^*] \qquad (j = 1, \ldots, n\,;\ k = 1, \ldots, m),$$

est notée $\frac{\partial}{\partial x} \mathrm{Sp}\,[axbx^*]$.

Désignons maintenant par q la matrice $n \times m$

$$q = \tilde{q} + \hat{q} \tag{28.42}$$

avec

$$\tilde{q} = a^* x b^*, \quad \hat{q} = axb. \tag{28.43}$$

Les éléments des matrices $n \times m$ $\tilde{\xi} = a^* x$ et $\hat{\xi} = ax$ sont

$$\tilde{\xi}_{j\mu} = \sum_{\rho=1}^{n} a_{\rho j} x_{\rho\mu}, \quad \hat{\xi}_{j\nu} = \sum_{l=1}^{n} a_{jl} x_{l\nu}.$$

Les éléments des matrices $n \times m$ $\tilde{q}$ et $\hat{q}$ sont

$$\tilde{q}_{jk} = \sum_{\mu=1}^{m} \tilde{\xi}_{j\mu} b_{k\mu}, \quad \hat{q}_{jk} = \sum_{\nu=1}^{m} \hat{\xi}_{j\nu} b_{\nu k}$$

ou

$$\tilde{q}_{jk} = \sum_{\mu=1}^{m} \sum_{\rho=1}^{n} a_{\rho j} x_{\rho\mu} b_{k\mu}, \quad \hat{q}_{jk} = \sum_{\nu=1}^{m} \sum_{l=1}^{n} a_{jl} x_{l\nu} b_{\nu k}. \tag{28.44}$$

Conformément à (42) et (44),

$$q_{jk} = \sum_{\rho=1}^{n} \sum_{\mu=1}^{m} a_{\rho j} x_{\rho\mu} b_{k\mu} + \sum_{l=1}^{n} \sum_{\nu=1}^{m} a_{jl} x_{l\nu} b_{\nu k} \qquad (j = 1, \ldots, n\,;\ k = 1, \ldots, m). \tag{28.45}$$

Il résulte des expressions (41) et (45) que

$$\frac{\partial}{\partial x_{jk}} \mathrm{Sp}\,[axbx^*] = q_{jk} \qquad (j = 1, \ldots, n\,;\ k = 1, \ldots, m). \tag{28.46}$$

Par suite, on a la relation

$$\frac{\partial}{\partial x} \mathrm{Sp}\,[axbx^*] = a^* x b^* + axb. \tag{28.47}$$

2) Examinons maintenant la matrice $m \times m$

$$\lambda = ex,$$

où e est la matrice $m \times n$; x, la matrice $n \times m$. Puisque

$$\lambda_{\rho\rho} = \sum_{l=1}^{n} e_{\rho l} x_{l\rho}, \qquad \mathrm{Sp}\,[ex] = \sum_{\rho=1}^{m} \sum_{l=1}^{n} e_{\rho l} x_{l\rho},$$

il vient

$$\frac{\partial}{\partial x_{jk}} \mathrm{Sp}\,[ex] = e_{kj} \qquad (j = 1, \ldots, n\,;\; k = 1, \ldots, m).$$

Donc

$$\frac{\partial}{\partial x} \mathrm{Sp}\,[ex] = e^{*}. \tag{28.48}$$

3) Soient maintenant h la matrice $n \times m$; x, la matrice $n \times m$. Considérons la matrice $n \times n$

$$\gamma = hx^{*}.$$

Etant donné que

$$\gamma_{\rho\rho} = \sum_{l=1}^{m} h_{\rho l} x_{\rho l}, \quad \mathrm{Sp}\,[hx^{*}] = \sum_{\rho=1}^{n} \sum_{l=1}^{m} h_{\rho l} x_{\rho l},$$

on a

$$\frac{\partial}{\partial x_{jk}} \mathrm{Sp}\,[hx^{*}] = h_{jk} \qquad (j = 1, \ldots, n\,;\; k = 1, \ldots, m),$$

et par suite,

$$\frac{\partial}{\partial x} \mathrm{Sp}\,[hx^{*}] = h. \tag{28.49}$$

4) Etudions encore quelques identités. Soient c la matrice $n \times m$ et d la matrice $m \times n$. Désignons par φ et ψ les matrices

$$\varphi = cd, \quad \psi = dc.$$

Comme

$$\varphi_{jj} = \sum_{l=1}^{m} c_{jl} d_{lj}, \qquad \mathrm{Sp}\,[cd] = \sum_{j=1}^{n} \sum_{l=1}^{m} c_{jl} d_{lj},$$

$$\psi_{ll} = \sum_{j=1}^{n} d_{lj} c_{jl}, \qquad \mathrm{Sp}\,[dc] = \sum_{l=1}^{m} \sum_{j=1}^{n} c_{jl} d_{lj},$$

on a l'identité

$$\mathrm{Sp}\,[cd] = \mathrm{Sp}\,[dc]. \tag{28.50}$$

5) De façon analogue, soient ξ et ζ des matrices $n \times m$. Désignons par q et p les matrices

$$q = \xi\zeta^{*}, \qquad p = \zeta\xi^{*}.$$

Comme

$$q_{jj} = \sum_{l=1}^{m} \xi_{jl}\zeta_{jl}, \quad \mathrm{Sp}\,[\xi\zeta^{*}] = \sum_{j=1}^{n} \sum_{l=1}^{m} \xi_{jl}\zeta_{jl},$$

$$p_{jj} = \sum_{l=1}^{m} \zeta_{jl}\xi_{jl}, \quad \mathrm{Sp}\,[\zeta\xi^{*}] = \sum_{j=1}^{n} \sum_{l=1}^{m} \xi_{jl}\zeta_{jl},$$

on a l'identité suivante:

$$\mathrm{Sp}\,[\xi\zeta^*] = \mathrm{Sp}\,[\zeta\xi^*]. \tag{28.51}$$

On trouve aisément à l'aide des formules (47)-(51) la matrice $\frac{\partial \mathscr{H}}{\partial G}$ de type $n \times m$ et la matrice $\frac{\partial \mathscr{H}}{\partial \Sigma}$ de type $n \times n$. En vertu de (38),

$$\frac{\partial \mathscr{H}}{\partial G} = -P(t)\,\Sigma^*(t)\,C^*(t) - P^*(t)\,\Sigma(t)\,C^*(t) + \\ + P(t)\,G(t)\,R^*(t) + P^*(t)\,G(t)\,R(t), \tag{28.52}$$

$$\frac{\partial \mathscr{H}}{\partial \Sigma} = A^*(t)\,P(t) - C^*(t)\,G^*(t)\,P(t) + P(t)A(t) - P(t)G(t)C(t). \tag{28.53}$$

Les conditions nécessaires pour que les commandes $\tilde{g}_{jk}$ $(j = 1, \ldots, n;\ k = 1, \ldots, m)$ minimisent la fonction $\mathscr{H}$ s'écrivent

$$\frac{\partial \mathscr{H}}{\partial g_{jk}} = 0 \qquad (j = 1, \ldots, n;\ k = 1, \ldots, m). \tag{28.54}$$

Le système des relations scalaires (54) est équivalent à la relation matricielle

$$\frac{\partial \mathscr{H}}{\partial G} = 0. \tag{28.55}$$

Conformément à (52), la condition (55) devient

$$-P(t)\,\Sigma^*(t)\,C^*(t) - P^*(t)\,\Sigma(t)\,C^*(t) + \\ + P(t)\,G(t)\,R^*(t) + P^*(t)\,G(t)\,R(t) = 0. \tag{28.56}$$

D'après le principe du maximum de Pontriaguine, les variables auxiliaires p_{ij} $(i, j = 1, \ldots, n)$ vérifient le système d'équations différentielles

$$\frac{dp_{ij}}{dt} = -\frac{\partial \mathscr{H}}{\partial \sigma_{ij}} \qquad (i, j = 1, \ldots, n) \tag{28.57}$$

équivalent à l'équation différentielle matricielle

$$\frac{dP}{dt} = -\frac{\partial \mathscr{H}}{\partial \Sigma}. \tag{28.58}$$

En raison de (53), l'équation (58) s'écrit

$$\frac{dP}{dt} = -P(t)\,[A(t) - G(t)\,C(t)] - [A(t) - G(t)\,C(t)]^*\,P(t). \tag{28.59}$$

Conformément à (24), dans notre problème les valeurs frontières de $p_{ij}(T)$ $(i, j = 1, \ldots, n)$ des variables auxiliaires sont, d'une façon analogue à (24.68),

$$p_{ij}(T) = \frac{\partial}{\partial \sigma_{ij}(T)}\,\mathrm{Sp}\,[L(T)\,\Sigma(T)] \qquad (i, j = 1, \ldots, n), \tag{28.60}$$

ce qu'on peut représenter en écriture matricielle

$$P(T)=\frac{\partial}{\partial\Sigma(T)}\operatorname{Sp}[L(T)\Sigma(T)]. \tag{28.61}$$

D'après (48)

$$\frac{\partial}{\partial\Sigma(T)}\operatorname{Sp}[L(T)\Sigma(T)]=L^*(T). \tag{28.62}$$

Puisque $L(T)$ est une matrice symétrique définie positive, il vient $L(T) = L^*(T)$ et, par conséquent,

$$P(T) = L(T). \tag{28.63}$$

Donc, la matrice $P(T)$ de type $n \times n$ est symétrique définie positive.

En transposant les deux membres de l'équation (59), on obtient

$$\frac{dP^*}{dt}=-P^*(t)[A(t)-G(t)C(t)]-[A(t)-G(t)C(t)]^*P^*(t). \tag{28.64}$$

Etant donné que $P(T)$ est une matrice symétrique, il vient, d'après (63), pour l condition aux limites

$$P^*(T) = L(T). \tag{28.65}$$

La comparaison de l'équation (59) et de la condition aux limites (63) avec l'équation (64) et la condition aux limites (65) montre que les équations différentielles et les conditions aux limites vérifiés par $P(t)$ et $P^*(t)$ coïncident; par suite, en raison du théorème d'unicité,

$$P(t) = P^*(t), \tag{28.66}$$

c'est-à-dire $P(t)$ est une matrice symétrique. Montrons encore que $P(t)$ est une matrice définie positive. En désignant par $A_1(t)$ et $A_2(t)$ les matrices $n \times n$

$$\begin{aligned} A_1(t) &= -[A(t) - G(t)C(t)]^*, \\ A_2(t) &= -[A(t) - G(t)C(t)], \end{aligned} \tag{28.67}$$

l'équation matricielle (59) se ramène à la forme

$$\frac{dP}{dt}=A_1(t)P(t)+P(t)A_2(t). \tag{28.68}$$

Soit $M_1(t, t_0)$ la matrice $n \times n$ qui vérifie l'équation différentielle matricielle

$$\frac{d}{dt}M_1(t, t_0)=A_1(t)M_1(t, t_0), \quad t\in[t_0, T], \ M_1(t_0, t_0)=E. \tag{28.69}$$

Supposons que les éléments de la matrice $A_1(t)$, et, par suite, ceux de la matrice $A_2(t)$ soient des fonctions bornées, continues par morceaux sur l'intervalle $[t_0, T]$. Alors, comme on le sait, la solution de l'équation (69) peut se mettre sous la forme d'un *matrizant* [17], c'est-à-dire sous celle d'une série absolument et uniformément convergente

$$M_1(t, t_0)=E+\Gamma A_1+\Gamma A_1\Gamma A_1+\Gamma A_1\Gamma A_1\Gamma A_1+\ldots$$

où

$$\Gamma A_1 = \int_{t_0}^{t} A_1(\tau_1)\, d\tau_1, \quad \Gamma A_1 \Gamma A_1 = \int_{t_0}^{t} A_1(\tau_1) \int_{t_0}^{\tau_1} A_1(\tau_2)\, d\tau_2\, d\tau_1,$$

$$\Gamma A_1 \Gamma A_1 \Gamma A_1 = \int_{t_0}^{t} A_1(\tau_1) \int_{t_0}^{\tau_1} A_1(\tau_2) \int_{t_0}^{\tau_2} A_1(\tau_3)\, d\tau_3\, d\tau_2\, d\tau_1, \ \ldots$$

Le matrizant $M_1(t, t_0)$ jouit de la propriété [17]

$$\det M_1(t, t_0) = \exp \int_{t_0}^{t} \operatorname{Sp} A_1(\tau)\, d\tau \tag{28.70}$$

et donc c'est une matrice régulière.

Désignons par $M_2(t, t_0)$ la matrice $n \times n$ qui vérifie l'équation différentielle matricielle

$$\frac{d}{dt} M_2(t, t_0) = A_2^*(t)\, M_2(t, t_0), \quad t \in [t_0, T], \quad M_2(t_0, t_0) = E. \tag{28.71}$$

D'une façon analogue à (70),

$$\det M_2(t, t_0) = \exp \int_{t_0}^{t} \operatorname{Sp} A_2^*(\tau)\, d\tau, \tag{28.72}$$

c'est-à-dire la matrice $M_2(t, t_0)$ est une matrice régulière.

L'équation différentielle (68) a comme solution:

$$P(t) = M_1(t, t_0)\, P(t_0)\, M_2^*(t, t_0); \qquad t \in [t_0, T], \tag{28.73}$$

dont la validité se vérifie par simple substitution. (73) entraîne que

$$P(T) = M_1(T, t_0)\, P(t_0)\, M_2^*(T, t_0). \tag{28.74}$$

Puisque d'après (63) $P(T) = L(T)$, c'est-à-dire puisque $P(T)$ est une matrice symétrique définie positive, il vient

$$\det P(T) \neq 0,$$

et, en vertu de (70) et (72), la relation (74) implique que

$$\det P(t_0) \neq 0.$$

Il résulte alors de la relation (73) que

$$\det P(t) \neq 0, \qquad t \in [t_0, T]. \tag{28.75}$$

Comme d'après (67) $A_2^*(t) = A_1(t)$, en vertu des équations (69) et (71), $M_2(t, t_0) = M_1(t, t_0)$ et l'expression (73) devient

$$P(t) = M_1(t, t_0)\, P(t_0)\, M_1^*(t_1, t_0), \qquad t \in [t_0, T]. \tag{28.76}$$

Les expressions (76), (75) et (66) entraînent que la matrice $P(t)$, $t \in [t_0, T]$, qui vérifie l'équation différentielle (59) et la condition aux limites (63), est symétrique définie positive.

En retenant que $P(t)$, $\Sigma(t)$ et $R(t)$ sont des matrices symétriques, la condition nécessaire (56) minimisant la fonction $\mathscr{H}$ peut se mettre sous la forme

$$P(t)\,[-\Sigma(t)\,C^*(t) + G(t)\,R(t)] = 0. \tag{28.77}$$

Comme d'après (75) $P(t)$ est une matrice régulière, la relation (77) ne peut avoir lieu que si l'on observe la condition

$$-\Sigma(t)\,C^*(t) + G(t)\,R(t) = 0. \tag{28.78}$$

L'équation matricielle (78), où l'inconnue est la matrice $G(t)$, possède une solution unique

$$G(t) = \Sigma(t)\,C^*(t)\,R^{-1}(t). \tag{28.79}$$

La matrice $G(t)$ définie par l'expression (79) constitue précisément la *commande optimale* minimisant la fonctionnelle (24).

En remplaçant dans l'équation différentielle (37) $G(t)$ par sa valeur tirée de (79) on ramène cette équation à la forme

$$\frac{d\Sigma}{dt} = A(t)\,\Sigma + \Sigma A^*(t) - \Sigma C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma + B(t)\,Q(t)\,B^*(t). \tag{28.80}$$

$$\Sigma(t_0) = \Sigma_0. \tag{28.81}$$

Dans ses ouvrages [34, 35], Kalman a montré que si Σ_0 est une matrice définie non négative, l'équation (80) possède une seule solution qui existe pour tout $t \geqslant t_0$.

D'après (20) et (79), l'équation différentielle du filtre optimal est

$$\frac{d\rho}{dt} = [A(t) - \Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)]\,\rho + $$
$$+ \Sigma(t)\,C^*(t)\,R^{-1}(t)\,z(t), \qquad \rho(t_0) = \bar{x}_0, \tag{28.82}$$

où $\Sigma(t)$ est la solution de l'équation différentielle de Riccati (80) qui vérifie la condition initiale (81).

La solution $\rho(t)$ $(t_0 \leqslant t \leqslant T)$ de l'équation différentielle (82) est une estimation optimale non déplacée de l'état $x(t)$ du système (I).

Faisons encore la remarque suivante. L'équation (82) peut se mettre sous la forme

$$\frac{d\rho}{dt} = A(t)\,\rho + \Sigma(t)\,C^*(t)\,R^{-1}(t)\,[z(t) - C(t)\,\rho], \qquad \rho(t_0) = \bar{x}_0, \tag{28.83}$$

où, d'après (6) et (11),

$$z(t) - C(t)\,\rho = C(t)\,[x(t) - \rho(t)] + v(t) = C(t)\,e(t) + v(t). \tag{28.84}$$

Ainsi, le filtre optimal est un système à retour dont l'entrée est définie par la fonction $C(t)\,e(t)$ dépendant de l'écart et par la perturbation additive du type de bruit blanc $v(t)$.

L'équation différentielle (80) et l'expression (79) montrent que les matrices $\Sigma(t)$ et $G(t)$ ne dépendent pas de la valeur de T intervenant dans la fonctionnelle (23) de l'extrémité de l'intervalle de temps d'observation $[t_0, T]$. La valeur de T ne détermine que la matrice $P(t)$ définie par l'équation différentielle (59) et la condition aux limites (63). Mais pour définir la matrice $G(t)$ à partir de la condition (56) nous n'avions appliqué que la propriété de symétrie et de non-dégénérescence de cette matrice $P(t)$. Ceci permet de conclure que le filtre optimal (82) donne une estimation optimale non déplacée *uniforme par rapport à t* de l'état $x(t)$ du système (1).

Il s'ensuit de (79), (80) et (81) que le filtre optimal ne dépend pas de la forme concrète de la matrice $L(T)$. Toutefois, l'hypothèse suivant laquelle cette dernière est symétrique définie positive est d'un grand intérêt, puisque ce n'est que dans ces conditions, comme le montrent (63), (74) et (76), que la matrice $P(t)$ est définie positive. L'expression (79) de la matrice G s'obtient à partir de la condition $\partial\mathscr{H}/\partial G = 0$; elle définit la valeur extrémale unique de la fonction $\mathscr{H}$. La matrice symétrique $\partial^2\mathscr{H}/\partial G^2$ de type $nm \times nm$ et dont les éléments sont les dérivées partielles $\partial^2\mathscr{H}/\partial\tilde{g}_{jk}\partial\tilde{g}_{il}$ ($j, i = 1, \ldots, n$; $k, l = 1, \ldots, m$) peut être mise sous la forme d'une matrice partitionnée symétrique S de type $m \times m$, dont les éléments $s_{\mu\nu}$ sont de la forme $s_{\mu\nu} = 2P(t)\, r_{\mu\nu}(t)$, où $r_{\mu\nu}$ sont les éléments de la matrice R. On peut montrer que la matrice $\partial^2\mathscr{H}/\partial G^2$ ne sera définie positive que dans le cas où $P(t)$ est une matrice définie positive.

La démonstration de cette proposition appartient à A. Formalski. Soit $\Xi(t)$ la matrice $n \times n$ vérifiant la condition $\Xi^*(t)\, P(t)\,\Xi(t) = \hat{E}$, où $\hat{E}$ est la matrice unité $n \times n$. Il vient

$$\begin{bmatrix} \Xi^* & \ldots & 0 \\ \cdot & \cdot\;\cdot & \cdot \\ 0 & \ldots & \Xi^* \end{bmatrix} \times \begin{bmatrix} Pr_{11} & \ldots & Pr_{1m} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ Pr_{m1} & \ldots & Pr_{mm} \end{bmatrix} \times \begin{bmatrix} \Xi & \ldots & 0 \\ \cdot & \cdot\;\cdot & \cdot \\ 0 & \ldots & \Xi \end{bmatrix} = \begin{bmatrix} \hat{E}r_{11} & \ldots & \hat{E}r_{1m} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \hat{E}r_{m1} & \ldots & \hat{E}r_{mm} \end{bmatrix}.$$

En désignant par N la matrice $m \times m$ qui vérifie la condition $N^*RN = E$, où E est la matrice unité $m \times m$, on a

$$\begin{bmatrix} \hat{E}n_{11} & \ldots & \hat{E}n_{m1} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \hat{E}n_{1m} & \ldots & \hat{E}n_{mm} \end{bmatrix} \times \begin{bmatrix} \hat{E}r_{11} & \ldots & \hat{E}r_{1m} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \hat{E}r_{m1} & \ldots & \hat{E}r_{mm} \end{bmatrix} \times \begin{bmatrix} \hat{E}n_{11} & \ldots & \hat{E}n_{1m} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \hat{E}n_{m1} & \ldots & \hat{E}n_{mm} \end{bmatrix} =$$

$$= \begin{bmatrix} \hat{E} & \ldots & 0 \\ \cdot & \cdot\;\cdot & \cdot \\ 0 & \ldots & \hat{E} \end{bmatrix},$$

ce qui prouve notre proposition.

On voit de ce qui précède que pour la valeur (79) de la matrice G la fonction $\mathscr{H}$ ne possède de minimum que sous la condition d'une

matrice $P(t)$ définie positive. Ainsi, la matrice G définie par l'expression (79) minimise la fonction $\mathcal{H}$ et, par conséquent, ne vérifie les conditions d'optimalité nécessaires pour les fonctionnelles de la forme (23) qu'à la condition que $L(T)$ soit une matrice définie positive.

Notons que la commande (79) vérifie non seulement les conditions nécessaires d'optimalité mais encore les conditions suffisantes d'optimalité, ce que l'on peut montrer, de même qu'au § 24, en examinant pour le problème donné l'équation de Bellman analogue à l'équation (24.46).

Montrons d'une façon analogue à (24.42) comment, sous la condition initiale (81), la matrice $\Sigma(t)$, qui est la solution de l'équation différentielle de Riccati (80), peut être exprimée à l'aide de la matrice fondamentale des solutions du système d'équations différentielles linéaires vectorielles

$$\begin{aligned} \frac{d\xi}{dt} &= -A^*(t)\,\xi + C^*(t)\,R^{-1}(t)\,C(t)\,\mu, \\ \frac{d\mu}{dt} &= B(t)\,Q(t)\,B^*(t)\,\xi + A(t)\,\mu, \end{aligned} \tag{28.85}$$

où $\xi(t)$ et $\mu(t)$ sont les vecteurs de dimension n.

On voit sans peine que (85) entraîne la relation

$$\mu(t) = \Sigma(t)\,\xi(t), \tag{28.86}$$

où $\Sigma(t)$ est la solution de l'équation de Riccati (80). En effet, conformément à (86), la deuxième équation (85) se met sous la forme

$$\frac{d\Sigma}{dt}\,\xi(t) + \Sigma(t)\,\frac{d\xi}{dt} = B(t)\,Q(t)\,B^*(t)\,\xi(t) + A(t)\,\Sigma(t)\,\xi(t). \tag{28.87}$$

En portant dans l'équation (87) au lieu de $\frac{d\xi}{dt}$ le second membre de la première équation (85) et en tenant compte de (86), on obtient

$$\left[\frac{d\Sigma}{dt} - A(t)\,\Sigma(t) - \Sigma(t)\,A^*(t) + \Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma(t) - \right.$$
$$\left. - B(t)\,Q(t)\,B^*(t)\right]\xi(t) = 0. \tag{28.88}$$

La relation (88) est identiquement vérifiée en vertu de l'équation différentielle de Riccati (80), qui définit la matrice $\Sigma(t)$. C'est ce qui justifie la relation (86).

Désignons par $g(t)$ la matrice fondamentale des solutions du système d'équations différentielles vectorielles (85) et par $\Pi(t, t_0)$, la matrice $2n \times 2n$

$$\Pi(t, t_0) = g(t)\,g^{-1}(t_0). \tag{28.89}$$

La solution du système d'équations vectorielles (85) s'écrit

$$\begin{bmatrix} \xi(t) \\ \mu(t) \end{bmatrix} = \Pi(t, t_0) \begin{bmatrix} \xi(t_0) \\ \mu(t_0) \end{bmatrix}. \tag{28.90}$$

En faisant figurer la matrice $\Pi(t, t_0)$ de type $2n \times 2n$ sous la forme d'une matrice partitionnée dont les éléments sont les matrices $n \times n$

$$\Pi(t, t_0) = \begin{bmatrix} \Pi_{11}(t, t_0) & \Pi_{12}(t, t_0) \\ \Pi_{21}(t, t_0) & \Pi_{22}(t, t_0) \end{bmatrix}, \tag{28.91}$$

on obtient à partir de (90) les relations

$$\begin{aligned} \xi(t) &= \Pi_{11}(t, t_0)\,\xi(t_0) + \Pi_{12}(t, t_0)\,\mu(t_0), \\ \mu(t) &= \Pi_{21}(t, t_0)\,\xi(t_0) + \Pi_{22}(t, t_0)\,\mu(t_0). \end{aligned} \qquad (28.92)$$

Puisque d'après (86) et (81),

$$\mu(t_0) = \Sigma_0 \xi(t_0), \qquad (28.93)$$

on a, en vertu de (86) et (92),

$$\Sigma(t)\,[\Pi_{11}(t, t_0) + \Pi_{12}(t, t_0)\,\Sigma_0] = [\Pi_{21}(t, t_0) + \Pi_{22}(t, t_0)]\,\Sigma_0, \qquad (28.94)$$

d'où l'on tire que

$$\Sigma(t) = [\Pi_{21}(t, t_0) + \Pi_{22}(t, t_0)\,\Sigma_0]\,[\Pi_{11}(t, t_0) + \Pi_{12}(t, t_0)\,\Sigma_0]^{-1}. \qquad (28.95)$$

Considérons encore quelques propriétés du processus aléatoire *d'innovation*

$$\chi(t) = z(t) - C(t)\,\rho(t) \qquad (28.96)$$

qui fait partie de l'équation différentielle (83) du filtre optimal.

D'après (10), (11) et (6), l'expression (96) peut s'écrire

$$\chi(t) = C(t)\,[x(t) - \hat{x}(t)] + v(t) = C(t)\,e(t) + v(t). \qquad (28.97)$$

Conformément à (14) et (7),

$$\mathrm{M}\,[\chi(t)] = 0, \qquad (28.98)$$

et donc $\chi(t)$ est un processus aléatoire gaussien à moyenne nulle.

Cherchons la matrice de corrélation de $\chi(t)$. D'après (97),

$$\begin{aligned} \mathrm{M}\,[\chi(t)\,\chi^*(\tau)] &= \mathrm{M}\,\{[C(t)\,e(t) + v(t)]\,[C(\tau)\,e(\tau) + v(\tau)]^*\} = \\ &= C(t)\,\mathrm{M}\,[e(t)\,e^*(\tau)]\,C^*(\tau) + C(t)\,\mathrm{M}\,[e(t)\,v^*(\tau)] + \\ &\quad + \mathrm{M}\,[v(t)\,e^*(\tau)]\,C^*(\tau) + \mathrm{M}\,[v(t)\,v^*(\tau)]. \end{aligned} \qquad (28.99)$$

Introduisons la notation

$$\mathscr{C}(t, \tau) = \mathrm{M}\,[e(t)\,e^*(\tau)]. \qquad (28.100)$$

Puisque (25) conduit à

$$e(t) = \Psi(t, \tau)\,e(\tau) + \int_\tau^t \Psi(t, s)\,B(s)\,w(s)\,ds - \int_\tau^t \Psi(t, s)\,G(s)\,v(s)\,ds, \qquad (28.101)$$

$$\mathscr{C}(t, \tau) = \Psi(t, \tau)\,\mathrm{M}\,[e(\tau)\,e^*(\tau)] + \int_\tau^t \Psi(t, s)\,B(s)\,\mathrm{M}\,[w(s)\,e^*(\tau)]\,ds - \int_\tau^t \Psi(t, s)\,G(s)\,\mathrm{M}\,[v(s)\,e^*(\tau)]\,ds. \qquad (28.102)$$

Comme pour $s > \tau$,

$$\mathrm{M}\,[w\,(s)\,e^*\,(\tau)] = 0,\ \ \mathrm{M}\,[v\,(s)\,e^*\,(\tau)] = 0, \qquad (28.103)$$

l'expression (102) devient conformément à (21)

$$\mathscr{C}\,(t,\ \tau) = \Psi\,(t,\ \tau)\,\Sigma\,(\tau). \qquad (28.104)$$

D'une façon analogue à (102),

$$\mathrm{M}\,[e\,(t)\,v^*\,(\tau)] = \Psi\,(t,\ \tau)\,\mathrm{M}\,[e\,(\tau)\,v^*(\tau)] + \\ + \int_\tau^t \Psi\,(t,\ s)\,B\,(s)\,\mathrm{M}\,[w\,(s)\,v^*\,(\tau)]\,ds - \\ - \int_\tau^t \Psi\,(t,\ s)\,G\,(s)\,\mathrm{M}\,[v\,(s)\,v^*\,(\tau)]\,ds. \qquad (28.105)$$

D'après (36) et (79),

$$\mathrm{M}\,[e\,(\tau)\,v^*\,(\tau)] = -\frac{1}{2}\,G\,(\tau)\,R\,(\tau) = -\frac{1}{2}\,\Sigma\,(\tau)\,C^*\,(\tau). \qquad (28.106)$$

En vertu de (8) et (79),

$$\int_\tau^t \Psi\,(t,\ s)\,G\,(s)\,\mathrm{M}\,[v\,(s)\,v^*\,(\tau)]\,ds = \\ = \int_\tau^t \Psi\,(t,\ s)\,G\,(s)\,R\,(s)\,\delta\,(s-\tau)\,ds = \\ = \frac{1}{2}\,\Psi\,(t,\ \tau)\,G\,(\tau)\,R\,(\tau) = \frac{1}{2}\,\Psi\,(t,\ \tau)\,\Sigma\,(\tau)\,C^*\,(\tau). \qquad (28.107)$$

Les processus aléatoires $w\,(t)$ et $v\,(t)$ ne sont pas corrélés, et donc,

$$\mathrm{M}\,[w\,(s)\,v^*\,(\tau)] = 0. \qquad (28.108)$$

Il résulte de (106), (107) et (108) que l'expression (105) se met sous la forme

$$\mathrm{M}\,[e\,(t)\,v^*\,(\tau)] = -\Psi\,(t,\ \tau)\,\Sigma\,(\tau)\,C^*\,(\tau). \qquad (28.109)$$

Le troisième terme de (99) s'annule du fait qu'avec $t > \tau$

$$\mathrm{M}\,[v\,(t)\,e^*\,(\tau)] = 0. \qquad (28.110)$$

Ainsi, d'après (104), (100), (109) et (110), l'expression (99) s'écrit

$$\mathrm{M}\,[\chi\,(t)\,\chi^*\,(\tau)] = C\,(t)\,\Psi\,(t,\ \tau)\,\Sigma\,(\tau)\,C^*\,(\tau) - \\ - C\,(t)\,\Psi\,(t,\ \tau)\,\Sigma\,(\tau)\,C^*\,(\tau) + \\ + \mathrm{M}\,[v\,(t)\,v^*\,(\tau)] = \mathrm{M}\,[v\,(t)\,v^*\,(\tau)], \qquad (28.111)$$

et, par suite, on a avec (8)

$$\mathbf{M}\,[\chi(t)\,\chi^*(\tau)] = R(t)\,\delta(t-\tau), \tag{28.112}$$

c'est-à-dire le processus aléatoire $\chi(t)$ est un bruit blanc de même matrice de corrélation que le processus aléatoire $v(t)$ intervenant additivement dans le signal observé (6).

§ 29. Théorie globale du filtrage

1. Estimation non déplacée à variance minimale de l'erreur. Considérons le système décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = A(t)\,x, \tag{29.1}$$

où x est le vecteur de dimension n ; $A(t)$, la matrice $n \times n$. Introduisons les notations

$$\Phi(t, \tau) = \theta(t)\,\theta^{-1}(\tau), \tag{29.2}$$

où $\theta(t)$ est la matrice fondamentale des solutions de l'équation (1).

La solution de l'équation (1) est

$$x(t) = \theta(t)\,\theta^{-1}(t_0)\,x(t_0). \tag{29.3}$$

Soit $t \in [t_0, T]$. D'une façon analogue à (3), on peut écrire que

$$x(T) = \theta(T)\,\theta^{-1}(t)\,x(t). \tag{29.4}$$

En multipliant les deux membres de la relation (4) à gauche par la matrice $\theta(t)\,\theta^{-1}(T)$, on obtient

$$\theta(t)\,\theta^{-1}(T)\,x(T) = \theta(t)\,\theta^{-1}(T)\,\theta(T)\,\theta^{-1}(t)\,x(t)$$

ou

$$x(t) = \theta(t)\,\theta^{-1}(T)\,x(T). \tag{29.5}$$

En récrivant l'expression (5), on a

$$x(t) = \Phi(t, T)\,x(T). \tag{29.6}$$

Supposons que sur l'intervalle $t_0 \leqslant t \leqslant T$ on mesure les valeurs du vecteur fonction $z(t)$ de dimension m défini par la formule

$$z(t) = C(t)\,x(t) + v(t), \tag{29.7}$$

où $v(t)$ est un processus aléatoire à moyenne nulle. Sa matrice de corrélation est

$$\mathbf{M}\,[v(t)\,v^*(\tau)] = R(t)\,\delta(t-\tau), \tag{29.8}$$

où $R(t)$ est une matrice symétrique définie positive.

L'expression (7) peut s'écrire

$$z(t) = C(t)\,\Phi(t, T)\,x(T) + v(t), \quad t_0 \leqslant t \leqslant T. \tag{29.9}$$

Ici $x(T)$ est un état fixé mais inconnu du système.

On demande de trouver la meilleure estimation de cet état. Considérons d'abord le problème d'estimation de la variable scalaire

$$\Omega = p^* x(T), \tag{29.10}$$

où p est le vecteur de dimension n arbitraire mais connu. Désignons par $\hat{\Omega}$ l'estimation de Ω. En général, $\hat{\Omega}$ est une variable aléatoire du fait qu'elle se détermine d'après la fonction $z(t)$ fournie par l'observation et intervenant additivement dans la fonction aléatoire $v(t)$.

Exigeons: 1) que l'estimation $\hat{\Omega}$ soit non déplacée, c'est-à-dire que la condition

$$\mathrm{M}[\hat{\Omega}] = p^* x(T) = \Omega \tag{29.11}$$

soit remplie; 2) que la variance de l'erreur d'estimation soit minimisée

$$\mathrm{M}[(\hat{\Omega} - \Omega)^2] = \min. \tag{29.12}$$

D é f i n i t i o n. *Le système* (9) *s'appelle complètement observable si pour tout* t_0 *il existe une valeur* $T(t_0)$ *telle que pour un certain paramètre* Ω, *défini par l'expression* (10), *on peut trouver une estimation linéaire non déplacée* $\hat{\Omega}$ *qui serait une fonctionnelle fournie par l'observation* $z(t)$ *dans l'intervalle de temps* $t_0 \leqslant t \leqslant T$.

Appelons estimation linéaire $\hat{\Omega}$ l'estimation

$$\hat{\Omega} = \int_{t_0}^{T} s^*(t) z(t)\, dt, \tag{29.13}$$

où $s(t)$ est un vecteur fonction de t de dimension m (peut-être, continu par morceaux).

Constatons qu'il se peut que le système (9) soit complètement observable pour certains vecteurs p, mais non pour tous.

T h é o r è m e 1 (s u r l ' o b s e r v a b i l i t é e n p r é s e n c e d e s p e r t u r b a t i o n s a l é a t o i r e s). *Le système* (9) *est complètement observable si et seulement si la matrice symétrique* $n \times n$

$$\mathcal{M}(t_0, T) = \int_{t_0}^{T} \Phi^*(t, T)\, C^*(t)\, R^{-1}(t)\, C(t)\, \Phi(t, T)\, dt \tag{29.14}$$

est définie positive pour un certain $T > t_0$.

D é m o n s t r a t i o n. 1°. Si $\mathcal{M}(t_0, T)$ est une matrice définie positive, il existe une inverse $\mathcal{M}^{-1}(t_0, T)$. En posant

$$s(t) = \hat{s}(t) = R^{-1}(t)\, C(t)\, \Phi(t, T)\, \mathcal{M}^{-1}(t_0, T)\, p, \quad t_0 \leqslant t \leqslant T \tag{29.15}$$

cherchons conformément à (13) l'estimation linéaire de la variable scalaire $\Omega = p^* x(T)$:

$$\hat{\Omega} = p^* \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) z(t)\, dt. \qquad (29.16)$$

Puisque $\mathbf{M}[v(t)] = 0$, d'après (9)

$$\mathbf{M}[z(t)] = \mathbf{M}[C(t) \Phi(t, T) x(T) + v(t)] = C(t) \Phi(t, T) x(T) \qquad (29.17)$$

et, en vertu de (16), (14) et (10),

$$\mathbf{M}[\hat{\Omega}] = p^* \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) C(t) \Phi(t, T)\, dt\, x(T) = p^* x(T) = \Omega. \qquad (29.18)$$

Ainsi, la condition (11) d'estimation non déplacée est respectée.

2°. Supposons que le système (9) soit complètement observable mais $\mathcal{M}(t_0, T)$ soit une matrice singulière. Il existe alors un vecteur $p \neq 0$ tel que la forme quadratique

$$\langle p, \mathcal{M}(t_0, T)\, p \rangle = 0. \qquad (29.19)$$

Soit

$$\zeta(t) = C(t) \Phi(t, T)\, p, \quad t \in [t_0, T]. \qquad (29.20)$$

(19) et (9) entraînent que la fonction $\zeta(t)$ coïncide identiquement avec la fonction $z(t)$ sur l'intervalle $[t_0, T]$ dans le cas où $x(T) = p$, et $v(t) \equiv 0$ pour $t \in [t_0, T]$.

Puisque d'après (19)

$$\int_{t_0}^{T} \zeta^*(t) R^{-1}(t) \zeta(t)\, dt = p^* \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) C(t) \Phi(t, T)\, dt\, p = p^* \mathcal{M}(t_0, T)\, p = \langle p, \mathcal{M}(t_0, T)\, p \rangle = 0 \qquad (29.21)$$

et $R^{-1}(t)$ est une matrice définie positive, il résulte de (21) que

$$\zeta(t) \equiv 0 \text{ pour } t \in [t_0, T]. \qquad (29.22)$$

Supposons maintenant que $\tilde{s}(t)$ détermine selon (13) l'estimation non déplacée de la variable scalaire $\Omega = p^* x(T)$. Pour $x(T) = p$ et $v(t) \equiv 0$, $t \in [t_0, T]$, d'après (13), (20), (22) et (11), on a

$$0 = \int_{t_0}^{T} \tilde{s}^*(t) \zeta(t)\, dt = \int_{t_0}^{T} \tilde{s}^*(t) C(t) \Phi(t, T)\, dt\, p = \hat{\Omega} = \Omega = \langle p, p \rangle. \qquad (29.23)$$

La relation obtenue contredit l'hypothèse initiale suivant laquelle $p \neq 0$. Cette réduction à l'absurde résulte de l'hypothèse que dans un système complètement observable la matrice $\mathcal{M}(t_0, T)$ puisse être singulière. Par suite, si le système (9) est complètement observable, la matrice $\mathcal{M}(t_0, T)$ est une matrice régulière. Etant donné que $\mathcal{M}(t_0, T)$ est une matrice de Gram et que $\det \mathcal{M}(t_0, T) \neq 0$, cette matrice est définie positive. Le théorème est démontré.

Cherchons maintenant l'équation différentielle vérifiée par la matrice $\mathcal{M}(t_0, T)$.

Notons d'abord que puisque

$$\theta^{-1}(T)\,\theta(T) = E,$$

où E est la matrice unité, il vient

$$\frac{d\theta^{-1}}{dT}\theta + \theta^{-1}\frac{d\theta}{dT} = 0,$$

d'où l'on tire que

$$\frac{d\theta^{-1}}{dT} = -\theta^{-1}\frac{d\theta}{dT}\theta^{-1}. \qquad (29.24)$$

Comme $\theta(t)$ est la matrice fondamentale de l'équation différentielle (1), on a

$$\frac{d\theta(T)}{dT} = A(T)\,\theta(T).$$

L'expression (24) devient ainsi

$$\frac{d\theta^{-1}(T)}{dT} = -\theta^{-1}(T)\,A(T).$$

Il s'ensuit que

$$\frac{d}{dT}\Phi(t, T) = \frac{d}{dT}[\theta(t)\,\theta^{-1}(T)] = -\theta(t)\,\theta^{-1}(T)\,A(T) =$$

$$= -\Phi(t, T)\,A(T), \quad \frac{d}{dT}\Phi^*(t, T) = -A^*(T)\,\Phi^*(t, T). \qquad (29.25)$$

En dérivant les deux membres de la relation (14), on obtient conformément à (25) l'équation différentielle vérifiée par la matrice $\mathcal{M}(t_0, T)$

$$\frac{d\mathcal{M}}{dT} = -A^*(T)\,\mathcal{M} - \mathcal{M}A(T) + C^*(T)\,R^{-1}(T)\,C(T). \qquad (29.26)$$

D'une façon analogue à (24)

$$\frac{d\mathcal{M}^{-1}}{dT} = -\mathcal{M}^{-1}\frac{d\mathcal{M}}{dT}\mathcal{M}^{-1}. \qquad (29.27)$$

D'après (26) et (27), l'équation différentielle vérifiée par la matrice $\mathcal{M}^{-1}(t_0, T)$ est

$$\frac{d\mathcal{M}^{-1}}{dT} = A(T)\,\mathcal{M}^{-1} + \mathcal{M}^{-1}A^*(T) - \mathcal{M}^{-1}C^*(T)\,R^{-1}(T)\,C(T)\,\mathcal{M}^{-1}. \qquad (29.28)$$

L'équation (28) coïncide avec l'équation différentielle associée à la matrice Σ, composée à partir de l'équation de Riccati (41.80) avec $Q(t) \equiv 0$. Ainsi, la matrice $\mathcal{M}^{-1}$ est analogue à la matrice Σ.

On voit facilement que la matrice $\mathcal{M}^{-1}(t_0, T)$ est une matrice de corrélation de l'erreur d'estimation linéaire non déplacée $\hat{x}(T)$ qui s'écrit d'une façon analogue à (16)

$$\hat{x}(T) = \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) z(t)\, dt. \tag{29.29}$$

En effet, d'après (17),

$$\mathrm{M}[z(t)] = C(t) \Phi(t, T) x(T)$$

et donc

$$\mathrm{M}[\hat{x}(T)] =$$
$$= \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) C(t) \Phi(t, T)\, dt\, x(T) = x(T). \tag{29.30}$$

Comme (29), (9) et (30) conduisent à

$$\hat{x}(T) - x(T) = \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) v(t)\, dt, \tag{29.31}$$

il vient

$$\mathrm{M}\{[\hat{x}(T) - x(T)][\hat{x}(T) - x(T)]^*\} =$$
$$= \mathrm{M}\{\mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) v(t)\, dt \times$$
$$\times \int_{t_0}^{T} v^*(\tau) R^{-1}(\tau) C(\tau) \Phi(\tau, T)\, d\tau \mathcal{M}^{-1}(t_0, T)\} =$$
$$= \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \int_{t_0}^{T} \Phi^*(t, T) C^*(t) R^{-1}(t) \times$$
$$\times \mathrm{M}[v(t) v^*(\tau)] R^{-1}(\tau) C(\tau) \Phi(\tau, T)\, d\tau\, dt \mathcal{M}^{-1}(t_0, T). \tag{29.32}$$

D'après (8)

$$\mathrm{M}[v(t) v^*(\tau)] = R(t) \delta(t - \tau)$$

et l'expression (32) devient

$$\mathbf{M}\{[\hat{x}(T)-x(T)][\hat{x}(T)-x(T)]^*\}=$$

$$=\mathcal{M}^{-1}(t_0, T)\int_{t_0}^{T}\Phi^*(t, T)C^*(t)R^{-1}(t)C(t)\Phi(t, T)\,dt\,\mathcal{M}^{-1}(t_0, T)=$$

$$=\mathcal{M}^{-1}(t_0, T), \qquad (29.33)$$

ce qui confirme la proposition énoncée dans ce qui précède.

Dans le cas où $v(t)$ est un processus aléatoire gaussien, la matrice $\mathcal{M}(t_0, T)$ coïncide avec la matrice d'information de Fischer.

Théorème 2. *Supposons que le système* (9) *soit complètement observable. Soit* $v(t)$ *un bruit blanc à moyenne nulle (non forcément gaussien) et à matrice de corrélation* $\mathbf{M}[v(t)v^*(\tau)] = R(t)\,\delta(t-\tau)$, *où* $R(t)$ *est la matrice symétrique définie positive. Dans ce cas l'estimation linéaire non déplacée à variance minimale de l'erreur, pour toute valeur scalaire* $\Omega = p^*x(T)$ *est* $\hat{\Omega} = p^*\hat{x}(T)$, où $\hat{x}(T)$ *est déterminée par la relation* (29):

$$\hat{x}(T)=\mathcal{M}^{-1}(t_0, T)\int_{t_0}^{T}\Phi^*(t, T)C^*(t)R^{-1}(t)z(t)\,dt.$$

Démonstration. 1°. Comme nous l'avons montré plus haut (30), $\mathbf{M}[\hat{x}(T)] = x(T)$. Par conséquent, $\mathbf{M}[\hat{\Omega}] = \Omega$.

2°. Soit σ l'estimation linéaire non déplacée de la variable scalaire $\Omega = p^*x(T)$ déterminée par la relation

$$\sigma=\int_{t_0}^{T}r^*(t)z(t)\,dt. \qquad (29.34)$$

Conformément à (9) l'expression (34) peut s'écrire

$$\sigma=\int_{t_0}^{T}r^*(t)[C(t)\Phi(t, T)x(T)+v(t)]\,dt. \qquad (29.35)$$

Puisque

$$\mathbf{M}[v(t)] = 0,$$

on tire de (35) que

$$\mathbf{M}[\sigma]=\int_{t_0}^{T}r^*(t)C(t)\Phi(t, T)x(T)\,dt. \qquad (29.36)$$

Il s'ensuit de (35) et (36) que l'erreur d'estimation

$$\varepsilon=\sigma-\mathbf{M}[\sigma]=\int_{t_0}^{T}r^*(t)v(t)\,dt. \qquad (29.37)$$

On tire de (37) que

$$\varepsilon^2 = \int_{t_0}^{T} \langle r(t),\ v(t)\rangle\, dt \int_{t_0}^{T} \langle v(\tau),\ r(\tau)\rangle\, d\tau =$$

$$= \int_{t_0}^{T}\int_{t_0}^{T} r^*(t)\, v(t)\, v^*(\tau)\, r(\tau)\, d\tau\, dt. \qquad (29.38)$$

En vertu de (38) on a la variance de l'erreur d'estimation

$$\mathrm{M}[\varepsilon^2] = \int_{t_0}^{T}\int_{t_0}^{T} r^*(t)\,\mathrm{M}[v(t)\,v^*(\tau)]\, r(\tau)\, d\tau\, dt =$$

$$= \int_{t_0}^{T}\int_{t_0}^{T} r^*(t)\, R(t)\,\delta(t-\tau)\, r(\tau)\, d\tau\, dt =$$

$$= \int_{t_0}^{T} r^*(t)\, R(t)\, r(t)\, dt = \int_{t_0}^{T} \langle r(t),\ R(t)\, r(t)\rangle\, dt. \qquad (29.39)$$

Si l'on tient compte de l'identité

$$r^*Rr = [\hat{s}^* + (r^* - \hat{s}^*)]\, R\, [\hat{s} + (r - \hat{s})] =$$

$$= 2\langle r,\ R\hat{s}\rangle - \langle \hat{s},\ R\hat{s}\rangle + (r - \hat{s})^* R\, (r - \hat{s}),$$

l'expression (39) se met sous la forme

$$\mathrm{M}[\varepsilon^2] = 2\int_{t_0}^{T} \langle r(t),\ R(t)\,\hat{s}(t)\rangle\, dt - \int_{t_0}^{T} \langle \hat{s}(t),\ R(t)\,\hat{s}(t)\rangle\, dt +$$

$$+ \int_{t_0}^{T} [r(t) - \hat{s}(t)]^* R(t)\, [r(t) - \hat{s}(t)]\, dt. \qquad (29.40)$$

Considérons maintenant certains calculs auxiliaires. D'après (15) et (14),

$$\int_{t_0}^{T} \hat{s}^*(t)\, R(t)\, \hat{s}(t)\, dt = p^* \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T)\, C^*(t)\, R^{-1}(t) \times$$

$$\times C(t)\, \Phi(t, T)\, dt\, \mathcal{M}^{-1}(t_0, T)\, p = p^* \mathcal{M}^{-1}(t_0, T)\, p = \langle p,\ \mathcal{M}^{-1}(t_0, T)\, p\rangle. \qquad (29.41)$$

Comme σ est une estimation linéaire non déplacée, on a d'après (36)

$$\mathrm{M}[\sigma] = \int_{t_0}^{T} r^*(t)\, C(t)\, \Phi(t, T)\, x(T)\, dt = p^* x(T). \qquad (29.42)$$

La relation (42) a lieu pour *tout* vecteur $x(T)$. C'est pourquoi (42) entraîne que

$$\int_{t_0}^{T} r^*(t)\, C(t)\, \Phi(t, T)\, dt = p^*. \tag{29.43}$$

Il résulte de (15) et (43) que

$$\int_{t_0}^{T} r^*(t) R(t) \hat{s}(t)\, dt = \int_{t_0}^{T} r^*(t)\, C(t)\, \Phi(t, T)\, \mathcal{M}^{-1}(t_0, T)\, p\, dt =$$

$$= p^* \mathcal{M}^{-1}(t_0, T)\, p = \langle p, \mathcal{M}^{-1}(t_0, T)\, p\rangle. \tag{29.44}$$

Cherchons maintenant la variance de l'erreur d'estimation non déplacée $\hat{\Omega} = p^* \hat{x}(T)$ de la variable scalaire $\Omega = p^* x(T)$. Puisque $\mathrm{M}[\hat{\Omega}] = \Omega$, on a

$$\hat{\Omega} - \mathrm{M}[\hat{\Omega}] = \hat{\Omega} - \Omega = p^*[\hat{x}(T) - x(T)] = p^* \nu,$$

où

$$\nu = \hat{x}(T) - x(T).$$

Il en résulte que

$$\mathrm{M}[p^* \nu (p^* \nu)^*] = \mathrm{M}[p^* \nu \nu^* p] = p^* \mathrm{M}[\nu \nu^*] p.$$

Puisque d'après (33)

$$\mathrm{M}[\nu \nu^*] = \mathcal{M}^{-1}(t_0, T),$$

on a

$$\mathrm{M}[(\hat{\Omega} - \Omega)^2] = p^* \mathcal{M}^{-1}(t_0, T)\, p = \langle p, \mathcal{M}^{-1}(t_0, T)\, p\rangle. \tag{29.45}$$

D'après (44) et (41), l'expression (40) se met sous la forme

$$\mathrm{M}[\varepsilon^2] = \langle p, \mathcal{M}^{-1}(t_0, T)\, p\rangle + \int_{t_0}^{T} [r(t) - \hat{s}(t)]^* R(t)\, [r(t) - \hat{s}(t)]\, dt, \tag{29.46}$$

et donc

$$\mathrm{M}[\varepsilon^2] \geqslant \langle p, \mathcal{M}^{-1}(t_0, T)\, p\rangle; \tag{29.47}$$

de plus, l'égalité n'a lieu que pour $r(t) \equiv \hat{s}(t)$, $t \in [t_0, T]$, c'est-à-dire pour $\sigma = \hat{\Omega}$.

En tenant compte de (37) et (45), on peut récrire la relation (47)

$$\mathrm{M}\{[\sigma - \mathrm{M}(\sigma)]^2\} \geqslant \mathrm{M}[(\hat{\Omega} - \Omega)^2];$$

de plus, l'égalité n'a lieu que pour $\sigma = \hat{\Omega}$. Le théorème est démontré.

Théorème 3. *L'estimation non déplacée $\hat{x}(T)$ à variance minimale de l'erreur, définie par la relation* (29)

$$\hat{x}(T) = \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T)\, C^*(t)\, R^{-1}(t)\, z(t)\, dt$$

est l'état à l'instant $t = T$ du système décrit par l'équation différentielle vectorielle

$$\frac{d\hat{x}}{dt} = A(t)\hat{x} + \hat{K}(t)[z(t) - C(t)\hat{x}], \tag{29.48}$$

où

$$\hat{K}(t) = \mathcal{M}^{-1}(t_0, t) C^*(t) R^{-1}(t). \tag{29.49}$$

D é m o n s t r a t i o n. Désignons par $\mathcal{T}(t, \tau)$ la matrice

$$\mathcal{T}(t, \tau) = \vartheta(t)\vartheta^{-1}(\tau), \tag{29.50}$$

où $\vartheta(t)$ est la matrice fondamentale des solutions de l'équation différentielle

$$\frac{d\hat{x}}{dt} = [A(t) - \mathcal{M}^{-1}(t, t_0) C^*(t) R^{-1}(t) C(t)]\hat{x} \tag{29.51}$$

construite à partir de l'équation (48) pour $z(t) \equiv 0$. Montrons d'abord que

$$\mathcal{T}(T, t) = \mathcal{M}^{-1}(t_0, T)\Phi^*(t, T)\mathcal{M}(t_0, t), \quad t_0 \leqslant t \leqslant T. \tag{29.52}$$

A cet effet remarquons qu'en vertu de l'équation différentielle (51),

$$\frac{d}{dt}\mathcal{T}(T, t) = -\mathcal{T}(T, t)[A(t) - \mathcal{M}^{-1}(t_0, t) C^*(t) R^{-1}(t) C(t)]. \tag{29.53}$$

D'autre part, en dérivant par rapport à t le second membre de l'expression (52), on obtient

$$\frac{d}{dt}\mathcal{T}(T, t) = -\mathcal{M}^{-1}(t_0, T)\Phi^*(t, T)\mathcal{M}(t_0, t)[A(t) - \mathcal{M}^{-1}(t_0, t) C^*(t) R^{-1}(t) C(t)]. \tag{29.54}$$

En comparant les expressions (53) et (54), on voit aisément que la relation (52) est vraie.

Si l'on prend en considération l'inégalité (47) obtenue dans ce qui précède, pour démontrer le théorème il suffit de prouver

1° que

$$\hat{x}(T) = \int_{t_0}^{T} \mathcal{T}(T, t)\hat{K}(t) z(t)\, dt \tag{29.55}$$

est une estimation non déplacée de l'état du système $x(T)$, et

2° que la variance de l'erreur d'estimation vaut $\mathcal{M}^{-1}(t_0, T)$.

En portant dans (55) l'expression (9) pour $z(t)$, on a

$$\mathrm{M}[\hat{x}(T)] = \int_{t_0}^{t} \mathcal{T}(T, t)\hat{K}(t) C(t)\Phi(t, T)\, dt\, x(T). \tag{29.56}$$

D'après (49),

$$\int_{t_0}^{T} \mathcal{H}(T, t)\,\hat{K}(t)\,C(t)\,\Phi(t, T)\,dt =$$

$$= \int_{t_0}^{T} \mathcal{H}(T, t)\,\mathcal{M}^{-1}(t_0, t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Phi(t, T)\,dt =$$

$$= \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T)\,\mathcal{M}(t_0, t)\,\mathcal{M}^{-1}(t_0, t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Phi(t, T)\,dt =$$

$$= \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Phi(t, T)\,dt = E, \qquad (29.57)$$

où E est la matrice unité.
Ainsi, la relation (56) devient

$$\mathrm{M}\,[x(T)] = x(T), \qquad (29.58)$$

c'est-à-dire que la première proposition demandée est démontrée.
Cherchons maintenant la variance de l'erreur

$$\mathrm{M}\,\{[\hat{x}(T) - x(T)]\,[\hat{x}(T) - x(T)]^*\} =$$

$$= \mathrm{M}\left\{ \int_{t_0}^{T} \mathcal{H}(T, \tau)\,\hat{K}(t)\,v(t)\,dt \int_{t_0}^{T} v^*(\tau)\,\hat{K}(\tau)\,\mathcal{H}^*(T, \tau)\,d\tau \right\} =$$

$$= \int_{t_0}^{T} \int_{t_0}^{T} \mathcal{H}(T, t)\,\hat{K}(t)\,\mathrm{M}\,[v(t)\,v^*(\tau)]\,\hat{K}^*(\tau)\,\mathcal{H}^*(T, \tau)\,d\tau\,dt. \qquad (29.59)$$

Puisque

$$\mathrm{M}\,[v(t)\,v^*(\tau)] = R(t)\,\delta(t - \tau),$$

la relation (59) se met sous la forme

$$\mathrm{M}\,\{[\hat{x}(T) - x(T)]\,[\hat{x}(T) - x(T)]^*\} =$$

$$= \int_{t_0}^{t} \mathcal{H}(T, t)\,\hat{K}(t)\,R(t)\,\hat{K}^*(t)\,\mathcal{H}^*(T, t)\,dt =$$

$$= \mathcal{M}^{-1}(t_0, T) \int_{t_0}^{T} \Phi^*(t, T)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Phi(t, T)\,dt\,\mathcal{M}^{-1}(t_0, T) =$$

$$= \mathcal{M}^{-1}(t_0, T) \qquad (29.60)$$

et donc la deuxième proposition demandée est également prouvée.
Le théorème est ainsi démontré.

Exemple. Considérons le système régi par les équations différentielles

$$\frac{dx_1}{dt}=x_2, \qquad \frac{dx_2}{dt}=-x_1. \tag{29.61}$$

Le système d'équations différentielles scalaires (61) peut être remplacé par l'équation vectorielle

$$\frac{dx}{dt}=A(t)\,x, \tag{29.62}$$

avec

$$x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad A(t)=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{29.63}$$

La matrice fondamentale des solutions de (62) est

$$\theta(t)=\begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}. \tag{29.64}$$

La matrice inverse $\theta^{-1}(T)$ s'écrit

$$\theta^{-1}(T)=\begin{bmatrix} \cos T & -\sin T \\ \sin T & \cos T \end{bmatrix}. \tag{29.65}$$

D'après (64) et (65),

$$\Phi(t, T)=\theta(t)\,\theta^{-1}(T)=\begin{bmatrix} \cos(t-T) & \sin(t-T) \\ -\sin(t-T) & \cos(t-T) \end{bmatrix}. \tag{29.66}$$

En admettant que dans l'exemple concerné

$$C(t) = [1 \quad 0], \tag{29.67}$$

on obtient

$$C(t)\,\Phi(t, T) = [\cos(t-T) \qquad \sin(t-T)]. \tag{29.68}$$

D'après (9), le signal observé $z(t)$ est une fonction scalaire définie par l'expression

$$z(t) = \cos(t-T)\,x_1(T) + \sin(t-T)\,x_2(T) + v(t), \quad t_0 \leqslant t \leqslant T, \tag{29.69}$$

où $v(t)$ est un processus aléatoire unidimensionnel du type bruit blanc à moyenne nulle et à fonction de corrélation

$$\mathrm{M}\,[v(t)\,v(\tau)] = R(t)\,\delta(t-\tau).$$

Adoptons pour notre exemple que

$$R(t) = r = \text{const}, \tag{29.70}$$

où r est une grandeur positive.

D'après (14), la matrice $\mathcal{M}(t_0, T)$ est dans notre exemple

$$\mathcal{M}(t_0, T) = \frac{1}{4r}\begin{bmatrix} 2(T-t_0)+\sin 2(T-t_0) & -1+\cos 2(T-t_0) \\ -1+\cos 2(T-t_0) & 2(T-t_0)-\sin 2(T-t_0) \end{bmatrix}. \tag{29.71}$$

L'inverse $\mathcal{M}^{-1}(t_0, T)$ s'écrit

$$\mathcal{M}^{-1}(t_0, T) = \begin{bmatrix} \frac{r[2(T-t_0)-\sin 2(T-t_0)]}{(T-t_0)^2-\sin^2(T-t_0)} & \frac{r[1-\cos 2(T-t_0)]}{(T-t_0)^2-\sin^2(T-t_0)} \\ \frac{r[1-\cos 2(T-t_0)]}{(T-t_0)^2-\sin^2(T-t_0)} & \frac{r[2(T-t_0)+\sin 2(T-t_0)]}{(T-t_0)^2-\sin^2(T-t_0)} \end{bmatrix}. \quad (29.72)$$

Le vecteur fonction $\hat{K}(t)$, gain de l'organe (48) prévu pour assurer l'estimation optimale, s'écrit d'après (49)

$$\hat{K}(t) = \begin{bmatrix} \hat{K}_1(t) \\ \hat{K}_2(t) \end{bmatrix} = \begin{bmatrix} \frac{2(t-t_0)-\sin 2(t-t_0)}{(t-t_0)^2-\sin^2(t-t_0)} \\ \frac{2\sin^2(t-t_0)}{(t-t_0)^2-\sin^2(t-t_0)} \end{bmatrix}, \quad t_0 \leqslant t \leqslant T. \quad (29.73)$$

Constatons que (73) entraîne $\lim\limits_{t\to t_0} \hat{K}_1(t) = \infty$, $\lim\limits_{t\to t_0} \hat{K}_2(t) = \infty$. Pour $t = t^* = t_0+\pi$, on a, conformément à (73), $\hat{K}_1(t^*) = \frac{2}{\pi}$, $\hat{K}_2(t^*) = 0$.

2. Stabilité des filtres optimaux de Kalman-Bucy. Nous avons étudié ci-dessus le système

$$\frac{dx}{dt} = A(t)\,x + B(t)\,w(t), \qquad z(t) = C(t)\,x(t) + v(t). \quad (29.74)$$

Ici $w(t)$ et $v(t)$ sont des processus aléatoires à moyennes nulles et à fonctions de corrélation

$$\mathbf{M}[w(t)\,w^*(\tau)] = Q(t)\,\delta(t-\tau), \qquad \mathbf{M}[v(t)\,v^*(\tau)] = R(t)\,\delta(t-\tau), \quad (29.75)$$

où $Q(t)$ et $R(t)$ sont des matrices symétriques définies positives.

Considérons avec le système (74) le système

$$\frac{dx}{dt} = -A^*(t)\,x + C^*(t)\,\tilde{w}(t), \qquad z(t) = B^*(t)\,x(t) + \tilde{v}(t), \quad (29.76)$$

où $\tilde{w}(t)$ et $\tilde{v}(t)$ sont des processus aléatoires à moyennes nulles et à matrices de corrélation

$$\mathbf{M}[\tilde{v}(t)\,\tilde{v}^*(\tau)] = Q^{-1}(t)\,\delta(t-\tau), \qquad \mathbf{M}[\tilde{w}(t)\,\tilde{w}^*(\tau)] = R^{-1}(t)\,\delta(t-\tau). \quad (29.77)$$

Nous avons désigné par $\theta(t)$ la matrice fondamentale des solutions de l'équation

$$\frac{dx}{dt} = A(t)\,x. \quad (29.78)$$

Introduisons la notation $\nu(t)$ pour la matrice fondamentale des solutions de l'équation conjuguée

$$\frac{dy}{dt} = -A^*(t)\,y. \tag{29.79}$$

On sait que les matrices $\theta(t)$ et $\nu(t)$ vérifient la relation (18.53)

$$\theta^*(t)\,\nu(t) = \nu^*(t)\,\theta(t) = E,$$

où E est la matrice unité. On en tire que

$$\theta^*(t) = \nu^{-1}(t), \quad \nu^*(t) = \theta^{-1}(t). \tag{29.80}$$

L'expression (80) donne lieu aux relations

$$\begin{aligned} [\nu(t)\,\nu^{-1}(T)]^* &= [\nu^{-1}(T)]^*\nu^*(t) = \theta(T)\,\theta^{-1}(t) = \Phi(T, t), \\ \nu(t)\,\nu^{-1}(T) &= [\theta^{-1}(t)]^*\theta^*(T) = [\theta(T)\,\theta^{-1}(t)]^* = \Phi^*(T, t). \end{aligned} \tag{29.81}$$

Pour que le système (76) soit complètement observable, il faut et il suffit que la matrice, analogue à (14),

$$\mathcal{W}(t_0, T) = \int_{t_0}^{T} [\nu(t)\,\nu^{-1}(T)]^* B(t)\,Q(t)\,B^*(t)\,\nu(t)\,\nu^{-1}(T)\,dt,$$

qui d'après (81) peut être mise sous la forme

$$\mathcal{W}(t_0, T) = \int_{t_0}^{T} \Phi(T, t)\,B(t)\,Q(t)\,B^*(t)\,\Phi^*(T, t)\,dt, \tag{29.82}$$

soit définie positive.

D é f i n i t i o n. *Le système* (74) *s'appelle complètement commandable si le système conjugué* (76) *est complètement observable.*

T h é o r è m e 4. *Le système* (74) *est complètement commandable si et seulement si la matrice* $\mathcal{W}(t_0, T)$, *déterminée par l'expression* (82), *est une matrice définie positive.*

D é m o n s t r a t i o n. La matrice $\mathcal{W}(t_0, T)$ est analogue à la matrice (14) et son caractère défini positif est une condition nécessaire et suffisante pour que le système (76) soit complètement observable. Selon la définition ci-dessus, le système (74) est complètement commandable.

La démonstration de ce théorème peut également être immédiate. Admettons que dans l'équation

$$\frac{dx}{dt} = A(t)\,x + B(t)\,w(t), \qquad x(t_0) = 0 \tag{29.83}$$

$w(t)$ désigne la commande à choisir de façon qu'à l'instant $t = T$ le système se trouve à l'état

$$x(T) = \mathcal{K}, \tag{29.84}$$

où $\mathcal{K}$ est un vecteur de dimension n donné.

En adoptant la commande $w(t)$ sous la forme

$$w(t) = Q(t)\, B^*(t)\, \Phi^*(T, t)\, \mathcal{W}^{-1}(t_0, T)\, \mathcal{K}, \tag{29.85}$$

où la matrice $\mathcal{W}(t_0, T)$ est définie par l'expression (82), on a conformément à (83),

$$x(T) = \int_{t_0}^{T} \Phi(T, t)\, B(t)\, Q(t)\, B^*(t)\, \Phi^*(T, t)\, dt\, \mathcal{W}^{-1}(t_0, T)\, \mathcal{K} = \mathcal{K}. \tag{29.86}$$

De la sorte, si $\mathcal{W}(t_0, T)$ est une matrice définie positive, le système (74) est complètement commandable.

La démonstration de la proposition, suivant laquelle si le système (74) est complètement commandable $\mathcal{W}(t_0, T)$ est une matrice définie positive, est analogue à celle donnée au § 11, point 6, section 2°, p. 189. Le théorème est démontré.

Dans ce qui suit nous allons donner les définitions d'un système *uniformément complètement observable* et d'un système *uniformément complètement commandable*. Dans ces définitions figurent les matrices $\mathcal{M}(t-\sigma, t)$ et $\mathcal{W}(t-\sigma, t)$ qui, d'après (14) et (82), sont

$$\mathcal{M}(t-\sigma, t) = \int_{t-\sigma}^{t} \Phi^*(\tau, t)\, C^*(\tau)\, R^{-1}(\tau)\, C(\tau)\, \Phi(\tau, t)\, d\tau, \tag{29.87}$$

$$\mathcal{W}(t-\sigma, t) = \int_{t-\sigma}^{t} \Phi(t, \tau)\, B(\tau)\, Q(\tau)\, B^*(\tau)\, \Phi^*(t, \tau)\, d\tau. \tag{29.88}$$

D é f i n i t i o n. a) *Le système* (74)

$$\frac{dx}{dt} = A(t)\, x + B(t)\, w(t),$$
$$z(t) = C(t)\, x(t) + v(t).$$

est dit uniformément complètement observable s'il existe des constantes positives fixées σ, α, β *telles que*

$$0 < \alpha E \leqslant \mathcal{M}(t-\sigma, t) \leqslant \beta E \quad \textit{pour tout } t \geqslant t_0 + \sigma. \tag{29.89}$$

b) *Le système* (74) *est dit uniformément complètement commandable si*

$$0 < \alpha E \leqslant \mathcal{W}(t-\sigma, t) \leqslant \beta E \quad \textit{pour tout } t \geqslant t_0 + \sigma. \tag{29.90}$$

Pour simplifier l'écriture, on suppose que les mêmes constantes σ, α, β font partie des deux inégalités (89) et (90) *. Par E on désigne la matrice unité.

* Constatons que pour les matrices symétriques M et N les conditions $M > N$ et $M \geqslant N$ signifient que la matrice $M - N$ est respectivement définie positive ou définie non négative.

Remarque. Dans les relations (75) et (77) données précédemment nous avons supposé que $Q(t)$ est une matrice définie positive. On voit sans peine que cette hypothèse *ne diminue pas la généralité* du problème.

En effet, soit dans les équations initiales

$$\frac{dx}{dt} = A(t)\,x + \tilde{B}(t)\,\breve{w}(t), \qquad z(t) = C(t)\,x(t) + v(t), \tag{*}$$

$\breve{w}(t)$ le processus aléatoire vectoriel gaussien de dimension r à moyenne nulle et à matrice de corrélation

$$\mathrm{M}\,[\breve{w}(t)\,\breve{w}^*(\tau)] = \tilde{Q}(t)\,\delta(t-\tau) \tag{**}$$

avec $Q(t)$ la matrice symétrique définie non négative $r \times r$.

La matrice $\tilde{Q}(t)$ peut être mise sous la forme

$$\tilde{Q}(t) = \tilde{\mathscr{C}}(t)\,\tilde{\mathscr{C}}^*(t),$$

où $\tilde{\mathscr{C}}(t)$ est la matrice réelle *singulière* $r \times r$ de même rang que la matrice $\tilde{Q}(t)$.

En introduisant les notations

$$\breve{w}(t) = \tilde{\mathscr{C}}(t)\,\hat{w}(t), \qquad \hat{B}(t) = \tilde{B}(t)\,\tilde{\mathscr{C}}(t),$$

on trouve que

$$\tilde{B}(t)\,\breve{w}(t) = \tilde{B}(t)\,\tilde{\mathscr{C}}(t)\,\hat{w}(t) = \hat{B}(t)\,\hat{w}(t).$$

La matrice de corrélation du processus aléatoire $\hat{w}(t)$ est

$$\mathrm{M}\,[\hat{w}(t)\,\hat{w}^*(\tau)] = E\delta(t-\tau),$$

où E est la matrice unité. La justification de ce fait est immédiate

$$\begin{aligned}\mathrm{M}\,[\breve{w}(t)\,\breve{w}^*(\tau)] &= \mathrm{M}\,[\tilde{\mathscr{C}}(t)\,\hat{w}(t)\,\hat{w}^*(\tau)\,\tilde{\mathscr{C}}^*(\tau)] = \\ &= \tilde{\mathscr{C}}(t)\,\mathrm{M}\,[\hat{w}(t)\,\hat{w}^*(\tau)]\,\tilde{\mathscr{C}}^*(\tau) = \tilde{\mathscr{C}}(t)\,E\delta(t-\tau)\,\tilde{\mathscr{C}}^*(\tau) = \\ &= \tilde{\mathscr{C}}(t)\,\tilde{\mathscr{C}}^*(\tau)\,\delta(t-\tau) = \tilde{\mathscr{C}}(t)\,\tilde{\mathscr{C}}^*(t)\,\delta(t-\tau) = \tilde{Q}(t)\,\delta(t-\tau),\end{aligned}$$

ce qui coïncide avec l'expression (**).

Dans ces conditions, les équations (*) deviennent

$$\frac{dx}{dt} = A(t)\,x + \hat{B}(t)\,\hat{w}(t), \qquad z(t) = C(t)\,x(t) + v(t). \tag{***}$$

Désignons maintenant par $Q(t)$ la matrice symétrique *définie positive* $r \times r$. Cette matrice peut s'écrire

$$Q(t) = S(t)\,S^*(t),$$

où $S(t)$ est une matrice réelle *régulière* $r \times r$. En introduisant les notations

$$\hat{w}(t) = S^{-1}(t)\,w(t), \qquad B(t) = \hat{B}(t)\,S^{-1}(t),$$

on a

$$\hat{B}(t)\,w(t) = \hat{B}(t)\,S^{-1}(t)\,w(t) = B(t)\,w(t).$$

La matrice de corrélation du processus aléatoire $w(t)$ est

$$\begin{aligned} \mathbf{M}[w(t)\,w^*(\tau)] &= \mathbf{M}[S(t)\,\hat{w}(t)\,\hat{w}^*(\tau)\,S^*(\tau)] = \\ &= S(t)\,\mathbf{M}[\hat{w}(t)\,\hat{w}^*(\tau)]\,S^*(\tau) = \\ &= S(t)\,E\delta(t-\tau)\,S^*(\tau) = S(t)\,S^*(\tau)\,\delta(t-\tau) = \\ &= S(t)\,S^*(t)\,\delta(t-\tau) = Q(t)\,\delta(t-\tau). \end{aligned}$$

L'équation (***) acquiert maintenant la forme de (74):

$$\frac{dx}{dt} = A(t)\,x + B(t)\,w(t),$$

$$z(t) = C(t)\,x(t) + v(t).$$

La matrice (88) de commandabilité uniforme du système (74) s'écrit

$$\mathscr{W}(t-\sigma, t) = \int_{t-\sigma}^{t} \Phi(t, \tau)\,B(\tau)\,Q(\tau)\,B^*(\tau)\,\Phi^*(t, \tau)\,d\tau,$$

où $Q(t)$ est la matrice symétrique définie positive $r \times r$.

Comme

$$B(t) = \hat{B}(t)\,S^{-1}(t) = \widetilde{B}(t)\,\widetilde{\mathscr{C}}(t)\,S^{-1}(t),$$

$$Q(t) = S(t)\,S^*(t), \qquad \widetilde{Q}(t) = \widetilde{\mathscr{C}}(t)\,\widetilde{\mathscr{C}}^*(t),$$

il vient

$$\begin{aligned} B(t)\,Q(t)\,B^*(t) &= \\ &= \widetilde{B}(t)\,\widetilde{\mathscr{C}}(t)\,S^{-1}(t)\,S(t)\,S^*(t)\,[S^{-1}(t)]^*\,\widetilde{\mathscr{C}}^*(t)\,\widetilde{B}^*(t) = \\ &= \widetilde{B}(t)\,\widetilde{\mathscr{C}}(t)\,\widetilde{\mathscr{C}}^*(t)\,\widetilde{B}^*(t) = \hat{B}(t)\,\hat{B}^*(t) = \widetilde{B}(t)\,\widetilde{Q}(t)\,\widetilde{B}^*(t), \end{aligned}$$

et la matrice (88) de commandabilité uniforme du système (74) peut être ramenée à la forme

$$\mathscr{W}(t-\sigma, t) = \int_{t-\sigma}^{t} \Phi(t, \tau)\,\widetilde{B}(\tau)\,\widetilde{Q}(\tau)\,\widetilde{B}^*(\tau)\,\Phi^*(t, \tau)\,d\tau,$$

ce qui n'est rien d'autre que la matrice de commandabilité uniforme du système initial (*).

Ainsi, *sans restreindre la généralité du problème, on peut considérer que dans le système* (74) $Q(t)$ *est une matrice définie positive, la matrice* $B(t)$ *tenant évidemment compte des transformations résultant de la normalisation du processus aléatoire initial au processus à matrice de corrélation définie positive* $Q(t)$.

Propriétés des solutions de l'équation de Riccati. Nous avons obtenu précédemment l'équation matricielle de Riccati (28.80)

$$\frac{d\Sigma}{dt} = A(t)\,\Sigma + \Sigma A^*(t) - \Sigma C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma + \\ + B(t)\,Q(t)\,B^*(t), \quad \Sigma(t_0) = \Sigma_0.$$

Désignons par $\Sigma(t;\,\Sigma_0,\,t_0)$ la solution de cette équation sous la condition initiale $\Sigma(t_0) = \Sigma_0$. Supposons dans ce qui suit que non seulement $R(t)$, mais encore $Q(t)$ soit une matrice définie positive.

Soit Σ_0 une matrice définie non négative que nous allons noter $\Sigma_0 \geqslant 0$. Le lemme qui suit permet de définir les limites entre lesquelles sont comprises les solutions de l'équation (28.80), dont l'existence locale et l'unicité résultent du fait que le second membre de cette équation vérifie la condition de Lipschitz.

L e m m e 1. *La solution* $\Sigma(t;\ \Sigma_0,\ t_0)$ *de l'équation de Riccati* (28.80) *à condition initiale* $\Sigma(t_0) = \Sigma_0 \geqslant 0$ *vérifie pour tout* $t > t_0$ *les conditions*:

$$0 \leqslant \Sigma(t;\ \Sigma_0,\ t_0) \leqslant \Phi(t,\ t_0)\,\Sigma_0\Phi^*(t,\ t_0) + \int_{t_0}^{t} \Phi(t,\ \tau)\,B(\tau)\,Q(\tau)\,B^*(\tau)\,\Phi^*(t,\ \tau)\,d\tau = \Phi(t,\ t_0)\,\Sigma_0\Phi^*(t,\ t_0) + \mathcal{W}(t_0,\ t), \qquad (29.91)$$

où $\Phi(t,\ t_0) = \theta(t)\,\theta^{-1}(t_0)$, *et* $\theta(t)$ *est la matrice fondamentale des solutions de l'équation différentielle vectorielle* (78)

$$\frac{dx}{dt} = A(t)\,x.$$

D é m o n s t r a t i o n. L'équation différentielle de Riccati (28.80) peut s'écrire

$$\frac{d\Sigma}{dt} = [A(t) - \Sigma C^*(t)\,R^{-1}(t)\,C(t)]\,\Sigma + \Sigma\,[A^*(t) - C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma] + \Sigma C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma + B(t)\,Q(t)\,B^*(t). \qquad (29.92)$$

Introduisons les notations $\Psi(t,\ t_0) = \varkappa(t)\,\varkappa^{-1}(t_0)$, où $\varkappa(t)$ est la matrice fondamentale des solutions de l'équation différentielle

$$\frac{d\xi}{dt} = [A(t) - \Sigma C^*(t)\,R^{-1}(t)\,C(t)]\,\xi. \qquad (29.93)$$

L'équation différentielle matricielle (92) à condition initiale $\Sigma(t_0) = \Sigma_0$ est équivalente à l'équation intégrale matricielle

$$\Sigma(t;\ \Sigma_0,\ t_0) = \Psi(t,\ t_0)\,\Sigma_0\Psi^*(t,\ t_0) + \int_{t_0}^{t} \Psi(t,\ \tau)\,[B(\tau)\,Q(\tau)\,B^*(\tau) + \Sigma(\tau)\,C^*(\tau)\,R^{-1}(\tau)\,C(\tau)\,\Sigma(\tau)]\,\Psi^*(t,\ \tau)\,d\tau. \qquad (29.94)$$

Si $t - t_0$ est suffisamment petit, on tire de (94) que $\Sigma(t;\ \Sigma_0,\ t_0) \geqslant 0$. Par conséquent, la solution $\Sigma(t;\ \Sigma_0,\ t_0)$ existe.

On peut passer directement de l'équation différentielle matricielle (28.80) à condition initiale $\Sigma(t_0) = \Sigma_0$ à l'équation intégrale

matricielle

$$\Sigma(t;\ \Sigma_0,\ t_0)=\Phi(t,\ t_0)\,\Sigma_0\Phi^*(t,\ t_0)+\int_{t_0}^{t}\Phi(t,\ \tau)\,B(\tau)\,Q(\tau)\,B^*(\tau)\,\Phi^*(t,\ \tau)\,d\tau-$$
$$-\int_{t_0}^{t}\Phi(t,\ \tau)\,\Sigma(\tau)\,C^*(\tau)\,R^{-1}(\tau)\,C(\tau)\,\Sigma(\tau)\,\Phi^*(t,\ \tau)\,d\tau. \quad (29.95)$$

Le terme retranché au second membre de (95) étant une matrice de Gram, c'est-à-dire une matrice définie positive ou non négative, (95) implique que

$$\Sigma(t;\ \Sigma_0,\ t_0)\leqslant\Phi(t,\ t_0)\,\Sigma_0\Phi^*(t,\ t_0)+\int_{t_0}^{t}\Phi(t,\ \tau)\,B(\tau)\,Q(\tau)\,B^*(\tau)\,\Phi^*(t,\ \tau)\,d\tau. \quad (29.96)$$

La limite supérieure ainsi obtenue de la solution de l'équation (28.80) permet de déterminer pour cette équation la constante de Lipschitz sur tout intervalle fini, aussi grand soit-il. Donc, l'équation différentielle de Riccati (28.80) possède une solution globale unique. Le lemme est démontré.

Au § 28 nous avons prouvé que la solution de (28.80) peut s'exprimer à l'aide de la matrice fondamentale des solutions du système d'équations différentielles vectorielles (28.85)

$$\frac{d\xi}{dt}=-A^*(t)\,\xi+C^*(t)\,R^{-1}(t)\,C(t)\,\mu,$$
$$\frac{d\mu}{dt}=B(t)\,Q(t)\,B^*(t)\,\xi+A(t)\,\mu,$$

où $\xi(t)$ et $\mu(t)$ sont les vecteurs de dimension n. Les équations (28.85) entraînent également la relation (28.86)

$$\mu(t)=\Sigma(t)\,\xi(t),$$

où $\Sigma(t)$ est la solution de l'équation de Riccati (28.80).

La matrice fondamentale des solutions de (28.85) est notée au § 28 $g(t)$. Par $\Pi(t,\ t_0)$ on désigne la matrice $2n\times 2n$

$$\Pi(t,\ t_0)=g(t)\,g^{-1}(t_0).$$

La matrice $\Pi(t,\ t_0)$ peut être mise sous la forme d'une matrice partitionnée dont les éléments sont des matrices $n\times n$

$$\Pi(t,\ t_0)=\begin{bmatrix}\Pi_{11}(t,\ t_0) & \Pi_{12}(t,\ t_0)\\ \Pi_{21}(t,\ t_0) & \Pi_{22}(t,\ t_0)\end{bmatrix}.$$

De plus, comme nous l'avons montré au § 28, la solution de l'équation de Riccati (28.80) s'écrit

$$\Sigma(t;\ \Sigma_0,\ t_0) = \mathscr{L}_{\Sigma_0}(t,\ t_0)\, \mathscr{D}_{\Sigma_0}^{-1}(t,\ t_0), \tag{29.97}$$

où

$$\mathscr{L}_{\Sigma_0}(t,\ t_0) = \Pi_{21}(t,\ t_0) + \Pi_{22}(t,\ t_0)\,\Sigma_0,$$
$$\mathscr{D}_{\Sigma_0}(t,\ t_0) = \Pi_{11}(t,\ t_0) + \Pi_{12}(t,\ t_0)\,\Sigma_0. \tag{29.98}$$

L e m m e 2. *Si* Σ_0 *est une matrice définie positive* ($\Sigma_0 > 0$), *la solution de l'équation de Riccati* (28.80) $\Sigma(t;\ \Sigma_0,\ t_0)$ *est une matrice définie positive pour tout* $t > t_0$.

Mais si Σ_0 *est une matrice définie non négative* ($\Sigma_0 \geqslant 0$), *alors que le système* (74) *est complètement uniformément commandable, la solution de l'équation de Riccati* (28.80) $\Sigma(t;\ \Sigma_0,\ t_0)$ *est une matrice définie positive pour tout* $t \geqslant t_0 + \sigma$.

D é m o n s t r a t i o n. Montrons d'abord que la matrice $\mathscr{D}_{\Sigma_0}(t,\ t_0)$ est régulière pour tout $t \geqslant t_0$ et tout $\Sigma_0 \geqslant 0$. En effet, s'il n'en était pas ainsi, il existerait des $t_1 \geqslant t_0$, $\Sigma_1 \geqslant 0$ et un vecteur non nul η tel que

$$\mathscr{D}_{\Sigma_1}(t_1,\ t_0)\,\eta = 0. \tag{29.99}$$

Considérons maintenant les équations (28.85) pour chercher leurs solutions vérifiant les conditions initales

$$\xi(t_0) = \eta,\ \mu(t_0) = \Sigma_1\eta. \tag{29.100}$$

Ces conditions satisfont à la relation (28.86) donnée dans ce qui précède.

Conformément à (28.90) et (28.91), à l'instant t_1 les fonctions $\xi(t)$ et $\mu(t)$ ont les valeurs suivantes

$$\begin{bmatrix} \xi(t_1) \\ \mu(t_1) \end{bmatrix} = \begin{bmatrix} \Pi_{11}(t_1,\ t_0) & \Pi_{12}(t_1,\ t_0) \\ \Pi_{21}(t_1,\ t_0) & \Pi_{22}(t_1,\ t_0) \end{bmatrix} \begin{bmatrix} \eta \\ \Sigma_1\eta \end{bmatrix}. \tag{29.101}$$

Il résulte de (101) que

$$\xi(t_1) = [\Pi_{11}(t_1,\ t_0) + \Pi_{12}(t_1,\ t_0)\,\Sigma_1]\,\eta \tag{29.102}$$

ou, d'après (98) et (99),

$$\xi(t_1) = \mathscr{D}_{\Sigma_1}(t_1,\ t_0)\,\eta = 0. \tag{29.103}$$

De plus, d'après (28.86) on a

$$\mu(t_1) = \Sigma(t_1;\ \Sigma_1,\ t_0)\,\xi(t_1). \tag{29.104}$$

Multiplions la première équation (28.85) à droite par μ et la deuxième équation (28.85) à gauche par ξ^*. En additionnant les premiers et les seconds membres des équations nouvellement obtenues

et en les intégrant dans les limites de t_0 à t_1, on obtient la relation

$$\xi^*(t_1)\,\mu(t_1)=\xi^*(t_0)\,\mu(t_0)+\int_{t_0}^{t_1}\langle B^*(s)\,\xi(s),\ Q(s)\,B^*(s)\,\xi(s)\rangle\,ds+ \\ +\int_{t_0}^{t_1}\langle C(s)\,\mu(s),\ R^{-1}(s)\,C(s)\,\mu(s)\rangle\,ds \qquad (29.105)$$

ou, d'après (103) et (100),

$$\langle\eta,\ \Sigma_1\eta\rangle+\int_{t_0}^{t}\langle B^*(s)\,\xi(s),\ Q(s)\,B^*(s)\,\xi(s)\rangle\,ds+ \\ +\int_{t_0}^{t}\langle C(s)\,\mu(s),\ R^{-1}(s)\,C(s)\,\mu(s)\rangle\,ds=0. \qquad (29.106)$$

Dans le lemme concerné, il suffit de supposer que $Q(t)$ soit une matrice définie non négative et non pas forcément une matrice définie positive. Puisque Σ_1 est une matrice définie non négative et $R(t)$ une matrice définie positive, pour justifier les relations (106), il faut que

$$C(t)\,\mu(t)=0,\quad t\in(t_0,\ t_1). \qquad (29.107)$$

De cette façon, la première équation de (28.85) devient

$$\frac{d\xi}{dt}=-A^*(t)\,\xi,\qquad \xi(t_0)=\eta. \qquad (29.108)$$

En intégrant cette équation on obtient

$$0=\xi(t_1)=\Phi^*(t_0,\ t_1)\,\eta=[\theta(t_0)\,\theta^{-1}(t_1)]^*\eta, \qquad (29.109)$$

où $\theta(t)$ est la matrice fondamentale des solutions de l'équation (1)

$$\frac{dx}{dt}=A(t)\,x.$$

Mais comme $\theta(t)$ est une matrice régulière, il s'ensuit de (109) que $\eta=0$. Cette réduction à l'absurde prouve que $\mathscr{D}_{\Sigma_1}(t,\ t_0)$ est une matrice régulière pour tout $t\geqslant t_0$ et tout $\Sigma_0\geqslant 0$, et l'expression (97) définit réellement la matrice $\Sigma(t;\ \Sigma_0,\ t_0)$ pour tout $t\geqslant t_0$ et tout $\Sigma_0\geqslant 0$.

Soit maintenant Σ_0 une matrice définie positive ($\Sigma_0>0$). Désignons par $\hat{\eta}$ le vecteur de dimension n

$$\hat{\eta}=\mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta. \qquad (29.110)$$

Supposons qu'il existe un vecteur non nul η tel que

$$\langle\hat{\eta},\ \Sigma(t;\ \Sigma_0,\ t_0)\,\hat{\eta}\rangle=0, \qquad (29.111)$$

c'est-à-dire que la matrice $\Sigma\,(t;\ \Sigma_0,\ t_0)$ n'est pas définie positive pour tout t. En tenant compte de (97), l'expression (111) peut s'écrire

$$\langle \mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta,\ \Sigma\,(t;\ \Sigma_0,\ t_0)\,\mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta\rangle = \\ = \langle \mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta,\ \mathscr{L}_{\Sigma_0}(t,\ t_0)\,\eta\rangle = \langle \eta,\ \mathscr{D}^*_{\Sigma_0}(t,\ t_0)\,\mathscr{L}_{\Sigma_0}(t,\ t_0)\,\eta\rangle = 0. \tag{29.112}$$

Supposons que les conditions initiales du système d'équations (28.85) soient

$$\xi\,(t_0) = \eta,\ \mu\,(t_0) = \Sigma_0\eta. \tag{29.113}$$

Les conditions initiales (113) vérifient la condition nécessaire (28.86).

Sous les conditions initiales (113), on a d'une façon analogue à (103)

$$\xi\,(t) = \mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta, \\ \mu\,(t) = \Sigma\,(t;\ \Sigma_0,\ t_0)\,\mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta = \mathscr{L}_{\Sigma_0}(t,\ t_0)\,\eta. \tag{29.114}$$

On tire de (114) que

$$\langle \xi\,(t),\ \mu\,(t)\rangle = \langle \mathscr{D}_{\Sigma_0}(t,\ t_0)\,\eta,\ \mathscr{L}_{\Sigma_0}(t,\ t_0)\,\eta\rangle = \\ = \langle \eta,\ \mathscr{D}^*_{\Sigma_0}(t,\ t_0)\,\mathscr{L}_{\Sigma_0}(t,\ t_0)\,\eta\rangle. \tag{29.115}$$

D'une façon analogue à (105), on a la relation

$$\langle \xi\,(t),\ \mu\,(t)\rangle = \langle \xi\,(t_0),\ \mu\,(t_0)\rangle + \\ + \int_{t_0}^{t} \langle B^*(s)\,\xi\,(s),\ Q\,(s)\,B^*(s)\,\xi\,(s)\rangle\,ds + \\ + \int_{t_0}^{t} \langle C\,(s)\,\mu\,(s),\ R^{-1}(s)\,C\,(s)\,\mu\,(s)\rangle\,ds \tag{29.116}$$

ou, d'après (115), (112) et (113),

$$0 = \langle \eta,\ \Sigma_0\eta\rangle + \int_{t_0}^{t} \langle B^*(s)\,\xi\,(s),\ Q\,(s)\,B^*(s)\,\xi\,(s)\rangle\,ds + \\ + \int_{t_0}^{t} \langle C\,(s)\,\mu\,(s),\ R^{-1}(s)\,C\,(s)\,\mu\,(s)\rangle\,ds. \tag{29.117}$$

Etant donné que Σ_0 et $R\,(t)$ sont des matrices définies positives et $Q\,(t) \geqslant 0$, la relation (117) n'est possible que pour $\hat{\eta} = 0$. Il en résulte que la relation (111) n'est possible que pour $\hat{\eta} = 0$ et on aboutit ainsi à une contradiction.

De cette façon, si Σ_0 est une matrice définie positive, la matrice $\Sigma\,(t;\ \Sigma_0,\ t_0)$ est également une matrice définie positive pour tout $t > t_0$.

Soit maintenant Σ_0 la matrice définie non négative, alors que la matrice

$$\mathscr{W}(t_0, T) = \int_{t_0}^{T} \Phi(T, s) B(s) Q(s) B^*(s) \Phi^*(T, s)\, ds \qquad (29.118)$$

est définie positive.

Supposons qu'il existe un vecteur non nul η tel qu'il donne lieu à la relation (111)

$$\langle \hat{\eta}, \Sigma(t; \Sigma_0, t_0) \hat{\eta} \rangle = 0 \text{ pour } t \geq T.$$

Les conditions initiales (113) conduisent aux relations (116)

$$0 = \langle \eta, \Sigma_0 \eta \rangle + \int_{t_0}^{t} \langle B^*(s) \xi(s), Q(s) B^*(s) \xi(s) \rangle\, ds + \int_{t_0}^{t} \langle C(s) \mu(s), R^{-1}(s) C(s) \mu(s) \rangle\, ds.$$

R étant une matrice définie positive, pour que la relation (116) existe, il faut que

$$C(s) \mu(s) = 0, \; s \in (t_0, t). \qquad (29.119)$$

La première équation (28.85) devient

$$\frac{d\xi}{dt} = -A^*(t) \xi, \; \xi(t_0) = \eta; \qquad (29.120)$$

sa solution s'écrit

$$\xi(t) = \Phi^*(t_0, t) \eta = [\theta(t_0) \theta^{-1}(t)]^* \eta. \qquad (29.121)$$

Pour que la relation (116) ait lieu, outre (119) il faut également que

$$\int_{t_0}^{t} \langle B^*(s) \xi(s), Q(s) B^*(s) \xi(s) \rangle\, ds = 0. \qquad (29.122)$$

En portant dans (122) l'expression (121) de $\xi(s)$, on ramène (122) à la forme

$$\langle \eta, J\eta \rangle = 0, \qquad (29.123)$$

où

$$J = \int_{t_0}^{T} \Phi(t_0, s) B(s) Q(s) B^*(s) \Phi^*(t_0, s)\, ds. \qquad (29.124)$$

Comme

$$\Phi(T, t_0) J \Phi^*(T, t_0) = \mathscr{W}(t_0, T) \qquad (29.125)$$

$\mathscr{W}(t_0, T)$ étant par hypothèse une matrice définie positive, et $\Phi(t, t_0)$ une matrice régulière, la matrice J est également une matrice définie positive. C'est pourquoi la relation (123) implique que $\eta = 0$. La contradiction obtenue prouve qu'avec $t > T$ la matrice $\Sigma(t; \Sigma_0, t_0)$ est définie positive.

On voit sans peine de ce qui vient d'être dit que si le système (74) est uniformément complètement commandable, c'est-à-dire si les inégalités (90) ont lieu, alors $\Sigma(t; \Sigma_0, t_0)$ est une matrice définie positive pour tout $t \geqslant t_0 + \sigma$. Le lemme est démontré.

L e m m e 3. *Si le système* (74)

$$\frac{dq}{dt} = A(t)\,q + B(t)\,w(t),$$
$$\zeta(t) = C(t)\,q(t) + v(t)$$

est uniformément complètement observable et uniformément complètement commandable, et si $\Sigma(t_0) = \Sigma_0$ est une matrice définie non négative, pour tout $t \geqslant t_0 + \sigma$ la matrice $\Sigma(t; \Sigma_0, t_0)$ est uniformément bornée et vérifie la condition

$$\Sigma(t; \Sigma_0, t_0) \leqslant \mathscr{M}^{-1}(t-\sigma, t) + \mathscr{W}(t-\sigma, t) \quad \text{pour tout } t \geqslant t_0 + \sigma.$$

D é m o n s t r a t i o n. Dans ce qui précède nous avons construit pour le système (1)

$$\frac{dx}{dt} = A(t)\,x$$

le filtre (48)

$$\frac{d\hat{x}}{dt} = A(t)\,\hat{x} + \hat{K}(t)\,[z(t) - C(t)\,\hat{x}],$$

où $\hat{K}(t)$ est défini par l'expression (50)

$$\hat{K}(t) = \mathscr{M}^{-1}(t_0, t)\, C^*(t)\, R^{-1}(t).$$

Le filtre (48) donne une estimation non déplacée $\hat{x}(T)$ de l'état $x(T)$ du système (1), c'est-à-dire l'estimation vérifiant la condition (58); $\mathrm{M}[\hat{x}(T)] = x(T)$, avec une variance minimale de l'erreur. De plus, on montre que cette estimation est définie par la relation (55)

$$\hat{x}(T) = \int_{t_0}^{T} \mathscr{K}(T, t)\,\hat{K}(t)\,z(t)\,dt,$$

où la matrice $\mathscr{K}(T, t)$ est définie par la relation (52). La fonction $z(t)$ intervenant dans l'expression (55) est présentée sous la forme (9):

$$z(t) = C(t)\,\Phi(t, T)\,x(T) + v(t), \quad t_0 \leqslant t \leqslant T.$$

La variance de l'erreur d'estimation $\hat{x}(T)$ de l'état $x(T)$ du système (1), donnée par le filtre (48), est, d'après (60),

$$\operatorname{cov}[x(T)] = \mathrm{M}\,\{[\hat{x}(T) - x(T)]\,[\hat{x}(T)] - x(T)]^*\} = \\ = \mathcal{M}^{-1}(t_0,\ T).$$

Supposons maintenant qu'on applique le filtre (48) pour obtenir l'estimation $\hat{q}(T)$ de l'état $q(T)$ du système (74).

En introduisant les notations

$$\delta q(T) = \int_{t_0}^{T} \Phi(T,\ \nu)\, B(\nu)\, w(\nu)\, d\nu \qquad (29.126)$$

on peut mettre l'état $q(T)$ du système (74) sous la forme

$$q(T) = x(T) + \delta q(T), \qquad (29.127)$$

où $x(T)$ est l'état du système (1).

L'expression (9), appliquée au système (74), devient

$$\zeta(T) = C(t)\,\Phi(t,\ T)\,[x(T) + \delta q(T)] + v(t), \qquad t_0 \leqslant t \leqslant T,$$

ou

$$\zeta(t) = C(t)\,\Phi(t,\ T)\,x(T) + v(t) + \\ + C(t)\,\Phi(t,\ T) \int_{t_0}^{T} \Phi(T,\ \nu)\, B(\nu)\, w(\nu)\, d\nu = \\ = z(t) + C(t)\,\Phi(t,\ T) \int_{t_0}^{T} \Phi(T,\ \nu)\, B(\nu)\, w(\nu)\, d\nu. \qquad (29.128)$$

Conformément à (55), l'estimation $\hat{q}(T)$ de l'état $q(T)$ du système (74) est

$$\hat{q}(T) = \hat{x}(T) + \int_{t_0}^{T} \mathcal{H}(T,\ \tau)\, \hat{K}(\tau)\, C(\tau) \times \\ \times \Phi(\tau,\ T) \int_{t_0}^{T} \Phi(T,\ \nu)\, B(\nu)\, w(\nu)\, d\nu\, d\tau. \qquad (29.129)$$

Le premier et le deuxième terme de l'expression (129) sont non corrélés. Donc

$$\operatorname{cov}[\hat{q}(T)] = \operatorname{cov}[\hat{x}(T)] + \\ + \operatorname{cov}\left[\int_{t_0}^{T} \mathcal{H}(T, \tau)\hat{K}(\tau)C(\tau)\Phi(\tau, T)d\tau \times \right. \\ \left. \times \int_{t_0}^{T} \Phi(T, \nu)B(\nu)w(\nu)d\nu\right], \qquad (29.130)$$

où, d'après (60),

$$\operatorname{cov}[\hat{x}(T)] = \mathcal{M}^{-1}(t_0, T).$$

En tenant compte que d'après (52)

$$\mathcal{H}(T, \tau) = \mathcal{M}^{-1}(t_0, T)\Phi^*(\tau, T)\mathcal{M}(t_0, \tau), \qquad t_0 \leqslant \tau \leqslant T,$$

on peut, en vertu de (50), mettre l'expression (130) sous la forme

$$\operatorname{cov}[\hat{q}(T)] = \mathcal{M}^{-1}(t_0, T) + \mathcal{L}, \qquad (29.131)$$

avec

$$\mathcal{L} = \operatorname{cov}\left[\int_{t_0}^{T} \mathcal{M}^{-1}(t_0, T)\Phi^*(\tau, T)\mathcal{M}(t_0, \tau)\mathcal{M}^{-1}(t_0, \tau) \times \right. \\ \left. \times C^*(\tau)R^{-1}(\tau)C(\tau)\Phi(\tau, T)d\tau \int_{t_0}^{T} \Phi(T, \nu)B(\nu)w(\nu)d\nu\right]$$

ou

$$\mathcal{L} = \operatorname{cov}\left[\mathcal{M}^{-1}(t_0, T)\int_{t_0} \Phi^*(\tau, T)C^*(\tau)R^{-1}(\tau) \times \right. \\ \left. \times C(\tau)\Phi(\tau, T)d\tau \int_{t_0}^{T} \Phi(T, \nu)B(\nu)w(\nu)d\nu\right]. \qquad (29.132)$$

D'après (14), le deuxième facteur du second membre de (132) est la matrice $\mathcal{M}(t_0, T)$; l'expression (132) devient donc

$$\mathcal{L} = \operatorname{cov}\left[\int_{t_0}^{T} \Phi(T, \nu)B(\nu)w(\nu)d\nu\right] \qquad (29.133)$$

ou

$$\mathcal{L} = \mathrm{M}\left[\int_{t_0}^{T} \Phi(T, \nu)B(\nu)w(\nu)d\nu \int_{t_0}^{T} w^*(\rho)B^*(\rho)\Phi^*(T, \rho)d\rho\right] = \\ = \int_{t_0}^{T}\int_{t_0}^{T} \Phi(T, \nu)B(\nu)\mathrm{M}[w(\nu)w^*(\rho)]B^*(\rho)\Phi^*(T, \rho)d\nu\, d\rho. \qquad (29.134)$$

Puisque

$$M [w (\nu) w^* (\rho)] = \delta (\nu - \rho) Q (\nu),$$

l'expression (134), en vertu de (82), se met sous la forme

$$\mathcal{L} = \int_{t_0}^{T} \Phi (T, \nu) B (\nu) Q (\nu) B^* (\nu) \Phi^* (T, \nu) d\nu = \mathcal{W} (t_0, T). \qquad (29.135)$$

Ainsi, d'après (131) et (135), on a

$$\operatorname{cov} [\hat{q} (T)] = \mathcal{M}^{-1} (t_0, T) + \mathcal{W} (t_0, T). \qquad (29.136)$$

Le filtre (48) n'étant pas optimal pour le système (74), la variance $\Sigma (t)$ de l'erreur d'estimation de l'état $q (T)$ par le filtre (28.82) optimal pour le système (74) vérifie la condition

$$\Sigma (T; \Sigma (t_0), t_0) \leqslant \operatorname{cov} [\hat{q} (T)] \qquad (29.137)$$

ou, en vertu de (136),

$$\Sigma (T; \Sigma (t_0), t_0) \leqslant \mathcal{M}^{-1} (t_0, T) + \mathcal{W} (t_0, T). \qquad (29.138)$$

Remplaçons maintenant la limite supérieure T de la relation (138) par t, et sa limite inférieure t_0 par $t - \sigma$. On a alors

$$\Sigma (t; \Sigma (t - \sigma), t - \sigma) \leqslant \mathcal{M}^{-1} (t - \sigma, t) + \\ + \mathcal{W} (t - \sigma, t). \qquad (29.139)$$

Dans (137), la matrice symétrique définie non négative $\Sigma (t_0) = \Sigma_0$ est la valeur de la matrice Σ à l'instant t_0. Par suite, à l'instant $t - \sigma$ la matrice $\Sigma (t - \sigma)$ présente la valeur de la solution de l'équation de Riccati (28.80) sous la condition initiale $\Sigma (t_0) = \Sigma_0$, ce que nous allons noter

$$\Sigma (t - \sigma) = \Sigma (t - \sigma; \Sigma_0, t_0), \qquad t \geqslant t_0 + \sigma. \qquad (29.140)$$

Maintenant on justifie sans peine l'identité

$$\Sigma (t; \Sigma (t - \sigma), t - \sigma) = \Sigma (t; \Sigma_0, t_0), \qquad t \geqslant t_0 + \sigma. \qquad (29.141)$$

Ainsi, la relation (139) devient

$$\Sigma (t; \Sigma_0, t_0) \leqslant \mathcal{M}^{-1} (t - \sigma, t) + \mathcal{W} (t - \sigma, t), \\ t \geqslant t_0 + \sigma. \qquad (29.142)$$

Le lemme est démontré.

Dans le lemme 2 il a été démontré que si $\Sigma (t_0) = \Sigma_0$ est une matrice définie non négative, alors que le système (74) est complètement uniformément commandable, $\Sigma (t; \Sigma_0, t_0)$ est une matrice définie positive pour tout $t \geqslant t_0 + \sigma$.

Puisque

$$\frac{d\Sigma^{-1}}{dt} = - \Sigma^{-1} \frac{d\Sigma}{dt} \Sigma^{-1},$$

en vertu de (28.80), la matrice Σ^{-1} satisfait à l'équation différentielle

$$\frac{d\Sigma^{-1}}{dt} = -\Sigma^{-1}A(t) - A^*(t)\Sigma^{-1} + C^*(t)R^{-1}(t)C(t) - \\ - \Sigma^{-1}B(t)Q(t)B^*(t)\Sigma^{-1}. \quad (29.143)$$

Ici la matrice $\Sigma^{-1}(t_0+\sigma) = \Sigma^{-1}(t_0+\sigma;\ \Sigma_0,\ t_0)$ est définie positive.

Nous avons introduit précédemment le système (76) qu'on appelle conjugué du système (74). Désignons par $\Xi(t)$ la matrice de corrélation de l'erreur d'estimation optimale de l'état du système (76). La comparaison de (74) et de (76) permet d'établir la correspondance suivante entre les paramètres de ces systèmes:

Système (74)	Système (76)
$A(t)$	$-A^*(t)$
$B(t)$	$C^*(t)$
$C(t)$	$B^*(t)$
$Q(t)$	$R^{-1}(t)$
$R(t)$	$Q^{-1}(t)$
$\mathcal{M}(t_0, T)$	$\mathcal{W}(t_0, T)$
$\mathcal{W}(t_0, T)$	$\mathcal{M}(t_0, T)$
$\Sigma(t)$	$\Xi(t)$

Si on en tient compte, on voit que

$$\Xi(t) = \Sigma^{-1}(t) \quad (29.144)$$

parce que si on remplace les paramètres de l'équation différentielle de Riccati (28.80) par les paramètres correspondants du système conjugué (76), (28.80) se ramène à la forme (143).

L e m m e 4. *Si le système* (74) *est uniformément complètement commandable et uniformément complètement observable, et si* $\Sigma(t_0) = \Sigma_0$ *est une matrice définie positive, la matrice* $\Sigma(t; \Sigma_0, t_0)$ *vérifie la condition*

$$[\mathcal{W}^{-1}(t-\sigma, t) + \mathcal{M}(t-\sigma, t)]^{-1} \leqslant \Sigma(t; \Sigma_0, t_0) \quad \text{pour tout } t \geqslant t_0 + \sigma.$$

D é m o n s t r a t i o n. En appliquant au système conjugué (76) l'estimation (142) obtenue dans le lemme précédent, on obtient

$$\Sigma^{-1}(t; \Sigma_0, t_0) \leqslant \mathcal{W}^{-1}(t-\sigma, t) + \mathcal{M}(t-\sigma, t), \qquad t \geqslant t_0 + \sigma. \qquad (29.145)$$

Constatons que si deux matrices carrées régulières $\mathcal{P}_1$ et $\mathcal{P}_2$ donnent lieu à la relation

$$\mathcal{P}_1 \leqslant \mathcal{P}_2, \qquad (29.146)$$

ce qui signifie que $\mathcal{P}_2 - \mathcal{P}_1$ est une matrice définie positive ou définie non négative, les inverses vérifient la condition

$$\mathcal{P}_1^{-1} \geqslant \mathcal{P}_2^{-1}, \qquad (29.147)$$

c'est-à-dire $\mathcal{P}_1^{-1} - \mathcal{P}_2^{-1}$ est une matrice définie positive ou définie non négative.

Conformément à (147), la relation (145) peut se ramener à la forme

$$[\mathcal{W}^{-1}(t-\sigma, t) + \mathcal{M}(t-\sigma, t)]^{-1} \leqslant \Sigma(t; \Sigma_0, t_0), \qquad t \geqslant t_0 + \sigma. \qquad (29.148)$$

Le lemme est démontré.

T h é o r è m e 5. *Si le système* (74)

$$\frac{dx}{dt} = A(t)x + B(t)w(t),$$
$$z(t) = C(t)x(t) + v(t)$$

est uniformément complètement observable et uniformément complètement commandable, le filtre optimal (28.82)

$$\frac{d\rho}{dt} = A(t)\rho + G(t)[z(t) - C(t)\rho],$$
$$G(t) = \Sigma(t)C^*(t)R^{-1}(t)$$

est uniformément asymptotiquement stable, c'est-à-dire que la solution triviale de l'équation différentielle homogène

$$\frac{d\rho}{dt} = [A(t) - \Sigma(t)C^*(t)R^{-1}(t)C(t)]\rho \qquad (29.149)$$

composée à partir de (28.82) *avec* $z(t) \equiv 0$ *est uniformément asymptotiquement stable.*

D é m o n s t r a t i o n. L'équation différentielle conjuguée de (149) est de la forme

$$\frac{d\tilde{\xi}}{dt} = [-A^*(t) + C^*(t) R^{-1}(t) C(t) \Sigma(t)] \tilde{\xi}. \qquad (29.150)$$

Dans ce qui précède (93) nous avons désigné par $\varkappa(t)$ la matrice fondamentale des solutions de l'équation (149), et par $\Psi(t, \tau)$, la matrice

$$\Psi(t, \tau) = \varkappa(t) \varkappa^{-1}(\tau). \qquad (29.151)$$

Introduisons la notation $\tilde{\vartheta}(t)$ pour la matrice fondamentale des solutions de l'équation (150), et $\Lambda(t, \tau)$ pour la matrice

$$\Lambda(t, \tau) = \tilde{\vartheta}(t) \tilde{\vartheta}^{-1}(\tau). \qquad (29.152)$$

On sait que les matrices $\varkappa(t)$ et $\tilde{\vartheta}(t)$ vérifient la condition (18.53)

$$\varkappa^*(t) \tilde{\vartheta}(t) = \tilde{\vartheta}^*(t) \varkappa(t) = E,$$

où E est la matrice unité. On en déduit que

$$\varkappa^*(t) = \tilde{\vartheta}^{-1}(t), \quad \tilde{\vartheta}^*(t) = \varkappa^{-1}(t),$$

$$\tilde{\vartheta}(t) \tilde{\vartheta}^{-1}(\tau) = [\varkappa^{-1}(t)]^* \varkappa^*(\tau) = [\varkappa(\tau) \varkappa^{-1}(t)]^*.$$

Ainsi

$$\Lambda(t, \tau) = \Psi^*(\tau, t). \qquad (29.153)$$

On demande de montrer qu'il existe des constantes positives c_1 et c_2 telles que *

$$\| \Lambda(t, t + \sigma) \| \leqslant c_1 e^{-c_2 \sigma} \quad \text{pour tout } t. \qquad (29.154)$$

La relation (154) est équivalente à la relation

$$\| \Psi^*(t + \sigma, t) \| = \| \Psi(t + \sigma, t) \| \leqslant c_1 e^{-c_2 \sigma} \quad \text{pour tout } t, \qquad (29.155)$$

ce qui traduit précisément la stabilité asymptotique uniforme de la solution triviale de (149).

Pour établir la validité de la relation (154), adoptons comme fonction de Liapounov la fonction

$$V(\tilde{\xi}, t) = \tilde{\xi}^* \Sigma(t) \tilde{\xi}. \qquad (29.156)$$

---

* Désignons la norme de la matrice A par $\| A \|$. En désignant par $\| x \|$ la norme du vecteur x de dimension n de l'espace X de dimension n et en posant $\| x \| = \sqrt{x^* x}$, adoptons comme norme $\| A \|$ de la matrice A la grandeur

$$\| A \| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\| Ax \|}{\| x \|}.$$

Puisque, d'après le lemme 2, pour $\Sigma_0 \geqslant 0$ la matrice $\Sigma(t;\ \Sigma_0,\ t_0) > 0$ avec $t \geqslant t_0 + \sigma$, et pour $\Sigma_0 > 0$, $\Sigma(t;\ \Sigma_0,\ t_0) > 0$ avec $t \geqslant t_0$, d'après (142), (145), (89) et (90), $\Sigma(t)$ est bornée supérieurement et inférieurement au moins pour tout $t \geqslant t_1 = t_0 + 2\sigma$:

$$0 < \frac{\alpha}{1+\alpha\beta} E \leqslant \Sigma(t) \leqslant \frac{1+\alpha\beta}{\alpha} E. \tag{29.157}$$

De cette façon, la fonction $V(\tilde{\xi},\ t)$ est uniformément bornée supérieurement et inférieurement

$$\frac{\alpha}{1+\alpha\beta} \|\tilde{\xi}\|^2 \leqslant V(\tilde{\xi},\ t) \leqslant \frac{1+\alpha\beta}{\alpha} \|\tilde{\xi}\|^2$$
$$\text{pour tout } t \geqslant t_1 = t_0 + 2\sigma. \tag{29.158}$$

Cherchons maintenant la fonction $\frac{dV}{dt}$ à partir de l'équation différentielle (150). En introduisant la notation

$$\hat{A}(t) = -A^*(t) + C^*(t) R^{-1}(t) C(t) \Sigma(t) \tag{29.159}$$

l'équation (150) peut se mettre sous la forme

$$\frac{d\tilde{\xi}}{dt} = \hat{A}(t) \tilde{\xi}. \tag{29.160}$$

En vertu de (156),

$$\frac{dV}{dt} = \tilde{\xi}^* \left(\hat{A}^* \Sigma + \frac{d\Sigma}{dt} + \Sigma \hat{A} \right) \tilde{\xi}.$$

Si l'on porte dans cette expression la valeur de $\frac{d\Sigma}{dt}$, tirée de (28.80), on a

$$\frac{dV}{dt} = \tilde{\xi}^* [\Sigma(t) C^*(t) R^{-1}(t) C(t) \Sigma(t) + B(t) Q(t) B^*(t)] \tilde{\xi}. \tag{29.161}$$

La matrice entre crochets du second membre de (161) étant une matrice de Gram, la forme quadratique du second membre de (161) ne peut être négative et la décroissance de la fonction $V(\tilde{\xi},\ t)$ avec $t \to \infty$ est impossible.

Pour évaluer l'ordre de croissance de la fonction $V(\tilde{\xi},\ t)$, cherchons la limite inférieure de l'intégrale par rapport à t du second membre de l'expression (161). Soit

$$t \leqslant \tau \leqslant t + \sigma. \tag{29.162}$$

Conformément à (150) et (152),

$$\tilde{\xi}(t+\sigma) = \Lambda(t+\sigma,\ \tau) \tilde{\xi}(\tau). \tag{29.163}$$

Comme

$$[\Lambda(t+\sigma,\ \tau)]^{-1} = \Lambda(\tau,\ t+\sigma), \tag{29.164}$$

(163) entraîne que

$$\tilde{\xi}(\tau)=\Lambda(\tau,\ t+\sigma)\,\tilde{\xi}(t+\sigma). \tag{29.165}$$

Introduisons la notation

$$u(\tau)=R^{-1}(\tau)\,C(\tau)\,\Sigma(\tau)\,\tilde{\xi}(\tau)= \\ =R^{-1}(\tau)\,C(\tau)\,\Sigma(\tau)\,\Lambda(\tau,\ t+\sigma)\,\tilde{\xi}(t+\sigma). \tag{29.166}$$

D'après (166), l'équation différentielle (150) peut se mettre sous la forme

$$\frac{d\tilde{\xi}}{dt}=-A^*(t)\,\tilde{\xi}+C^*(t)\,u(t). \tag{29.167}$$

Nous avons désigné précédemment par $\nu(t)$ la matrice fondamentale des solutions de l'équation (79)

$$\frac{dy}{dt}=-A^*(t)\,y.$$

En introduisant la notation

$$h(t,\ \tau)=\nu(t)\,\nu^{-1}(\tau) \tag{29.168}$$

appliquons l'équation (167) pour obtenir

$$\tilde{\xi}(t+\sigma)=h(t+\sigma,\ \tau)\,\tilde{\xi}(\tau)+\int_{\tau}^{t+\sigma}h(t+\sigma,\ \nu)\,C^*(\nu)\,u(\nu)\,d\nu. \tag{29.169}$$

Comme

$$[h(t+\sigma,\ \tau)]^{-1}=h(\tau,\ t+\sigma),\qquad h(\tau,\ t+\sigma)\,h(t+\sigma,\ \nu)=h(\tau,\ \nu),$$

(169) implique que

$$h(\tau,\ t+\sigma)\,\tilde{\xi}(t+\sigma)=\tilde{\xi}(\tau)+\int_{\tau}^{t+\sigma}h(\tau,\ \nu)\,C^*(\nu)\,u(\nu)\,d\nu. \tag{29.170}$$

Puisque

$$h(t,\ \tau)=\Phi^*(\tau,\ t)$$

la relation (170) devient

$$\Phi^*(t+\sigma,\ \tau)\,\tilde{\xi}(t+\sigma)=\tilde{\xi}(\tau)+\int_{\tau}^{t+\sigma}\Phi^*(\nu,\ \tau)\,C^*(\nu)\,u(\nu)\,d\nu$$

ou

$$\tilde{\xi}(\tau)=\Phi^*(t+\sigma,\ \tau)\,\tilde{\xi}(t+\sigma)-\int_{\tau}^{t+\sigma}\Phi^*(\nu,\ \tau)\,C^*(\nu)\,u(\nu)\,d\nu. \tag{29.171}$$

D'après (165), l'expression (171) devient

$$\Lambda(\tau,\ t+\sigma)\tilde{\xi}(t+\sigma)=\Phi^*(t+\sigma,\ \tau)\tilde{\xi}(t+\sigma)-\\ -\int_{\tau}^{t+\sigma}\Phi^*(\nu,\ \tau)C^*(\nu)u(\nu)\,d\nu. \quad (29.172)$$

Introduisons les notations

$$\int_{t}^{t+\sigma} u^*(\nu)R(\nu)u(\nu)\,d\nu=\int_{t}^{t+\sigma}\tilde{\xi}^*(\nu)\Sigma(\nu)C^*(\nu)R^{-1}(\nu)\times\\ \times C(\nu)\Sigma(\nu)\tilde{\xi}(\nu)\,d\nu=\tilde{\xi}^*(t+\sigma)\int_{t}^{t+\sigma}\Lambda^*(\nu,\ t+\sigma)\times\\ \times\Sigma(\nu)C^*(\nu)R^{-1}(\nu)C(\nu)\Sigma(\nu)\Lambda(\nu,\ t+\sigma)\,d\nu\ \tilde{\xi}(t+\sigma)=\\ =\varepsilon^2\|\tilde{\xi}(t+\sigma)\|^2, \quad (29.173)$$

$$\int_{t}^{t+\sigma}\tilde{\xi}^*(\nu)B(\nu)Q(\nu)B^*(\nu)\tilde{\xi}(\nu)\,d\nu=\tilde{\xi}^*(t+\sigma)\int_{t}^{t+\sigma}\Lambda^*(\nu,\ t+\sigma)\times\\ \times B(\nu)Q(\nu)B^*(\nu)\Lambda(\nu,\ t+\sigma)\,d\nu\,\tilde{\xi}(t+\sigma)=\eta^2\|\tilde{\xi}(t+\sigma)\|^2 \quad (29.174)$$

où ε^2 et η^2 dépendent de $t+\sigma$.

La somme des fonctions sous le signe somme de (173) et (174) représente le second membre de la relation (161). C'est pourquoi en désignant

$$\delta^2=\varepsilon^2+\eta^2 \quad (29.175)$$

on trouve à partir de (161), (173) et (174) que

$$V(\tilde{\xi}(t+\sigma),\ t+\sigma)-V(\tilde{\xi}(t),\ t)=\\ =\delta^2\|\tilde{\xi}(t+\sigma)\|^2. \quad (29.176)$$

Passons maintenant à certaines transformations qui nous serviront pour évaluer la grandeur δ^2.

Remarquons au préalable que (171) peut s'écrire

$$\tilde{\xi}(\tau)=\Phi^*(t+\sigma,\ \tau)\Big[\tilde{\xi}(t+\sigma)-\int_{\tau}^{t+\sigma}\Phi^*(\nu,\ t+\sigma)C^*(\nu)u(\nu)\,d\nu \quad (29.177)$$

ou

$$\tilde{\xi}(\tau)=\Phi^*(t+\sigma,\ \tau)[\tilde{\xi}(t+\sigma)-S(\tau)], \quad (29.178)$$

avec

$$S(\tau)=\int_{\tau}^{t+\sigma}\Phi^*(\nu,\ t+\sigma)C^*(\nu)u(\nu)\,d\nu. \quad (29.179)$$

L'expression (174) se met sous la forme

$$\eta^2 \| \tilde{\xi}(t+\sigma) \|^2 = \int_t^{t+\sigma} [\tilde{\xi}^*(t+\sigma) - S^*(\tau)] \times \\ \times \Phi(t+\sigma, \tau) B(\tau) Q(\tau) B^*(\tau) \Phi^*(t+\sigma, \tau) \times \\ \times [\tilde{\xi}(t+\sigma) - S(\tau)] \, d\tau \quad (29.180)$$

ou

$$\eta^2 \| \tilde{\xi}(t+\sigma) \|^2 = \tilde{\xi}^*(t+\sigma) \int_t^{t+\sigma} \tilde{\mathscr{L}}(\tau) \, d\tau \, \tilde{\xi}(t+\sigma) - \\ - \tilde{\xi}^*(t+\sigma) \int_t^{t+\sigma} \tilde{\mathscr{L}}(\tau) S(\tau) \, d\tau - \int_t^{t+\sigma} S^*(\tau) \tilde{\mathscr{L}}(\tau) \, d\tau \times \\ \times \tilde{\xi}(t+\sigma) + \int_t^{t+\sigma} S^*(\tau) \tilde{\mathscr{L}}(\tau) S(\tau) \, d\tau, \quad (29.181)$$

avec

$$\tilde{\mathscr{L}}(\tau) = \Phi(t+\sigma, \tau) B(\tau) Q(\tau) B^*(\tau) \Phi^*(t+\sigma, \tau). \quad (29.182)$$

Puisque d'après (182), $\tilde{\mathscr{L}}(\tau)$ est une matrice de Gram, il vient

$$\int_t^{t+\sigma} S^*(\tau) \tilde{\mathscr{L}}(\tau) S(\tau) \, d\tau \geqslant 0. \quad (29.183)$$

Etant donné que $\tilde{\mathscr{L}}(\tau) = \tilde{\mathscr{L}}^*(\tau)$, le deuxième et le troisième terme du second membre de (181) coïncident.

En éliminant le dernier terme du second membre de (181) et en retenant que d'après (88)

$$\int_t^{t+\sigma} \tilde{\mathscr{L}}(\tau) \, d\tau = \mathscr{W}(t, t+\sigma),$$

on obtient la relation

$$\eta^2 \| \tilde{\xi}(t+\sigma) \|^2 \geqslant \langle \tilde{\xi}(t+\sigma), \mathscr{W}(t, t+\sigma) \tilde{\xi}(t+\sigma) \rangle - \\ - 2 \left\langle \tilde{\xi}(t+\sigma), \int_t^{t+\sigma} \tilde{\mathscr{L}}(\tau) S(\tau) \, d\tau \right\rangle. \quad (29.184)$$

Constatons que selon (90) on a

$$\mathscr{W}(t, t+\sigma) \geqslant \alpha E > 0$$

et, donc,

$$\langle \tilde{\xi}(t+\sigma),\ \mathscr{W}(t,\ t+\sigma)\,\tilde{\xi}(t+\sigma)\rangle \geqslant \alpha \,\|\,\tilde{\xi}(t+\sigma)\,\|^2. \quad (29.185)$$

C'est pourquoi (184) peut être remplacée par l'inégalité

$$\eta^2 \|\tilde{\xi}(t+\sigma)\|^2 \geqslant \alpha \|\tilde{\xi}(t+\sigma)\|^2 - 2\left\langle \tilde{\xi}(t+\sigma),\ \int_t^{t+\sigma} \tilde{\mathscr{L}}(\tau)\, S(\tau)\, d\tau \right\rangle. \quad (29.186)$$

Passons maintenant à l'estimation du deuxième terme du second membre de (186). Conformément à l'inégalité de Bouniakovski-Schwarz,

$$\|S(\tau)\|^2 = \left\| \int_\tau^{t+\sigma} \Phi^*(\nu,\ t+\sigma)\, C^*(\nu)\, u(\nu)\, d\nu \right\|^2 \leqslant$$

$$\leqslant \left[\int_t^{t+\sigma} \|\Phi^*(\nu,\ t+\sigma)\, C^*(\nu)\, R^{-\frac{1}{2}}(\nu)\| \, \|R^{\frac{1}{2}}(\nu)\, u(\nu)\| \, d\nu \right]^2 \leqslant$$

$$\leqslant \left\| \int_t^{t+\sigma} \Phi^*(\nu,\ t+\sigma)\, C^*(\nu)\, R^{-1}(\nu)\, C(\nu)\, \Phi(\nu,\ t+\sigma)\, d\nu \right\| \times$$

$$\times \left\| \int_t^{t+\sigma} u^*(\nu)\, R(\nu)\, u(\nu)\, d\nu \right\|. \quad (29.187)$$

D'après (87) et (173), (187) peut s'écrire

$$\|S(\tau)\|^2 \leqslant \varepsilon^2 \,\|\mathscr{M}(t,\ t+\sigma)\|\, \|\tilde{\xi}(t+\sigma)\|^2. \quad (29.188)$$

Puisque d'après (89), $\mathscr{M}(t,\ t+\sigma) \leqslant \beta E$, où E est la matrice unité, il vient

$$\|\mathscr{M}(t,\ t+\sigma)\| \leqslant \beta \quad (29.189)$$

et on a à partir de (188),

$$\|S(\tau)\| \leqslant \varepsilon\beta^{1/2} \|\tilde{\xi}(t+\sigma)\|. \quad (29.190)$$

En vertu de (190),

$$\left\| \int_t^{t+\sigma} \tilde{\mathscr{L}}(\tau)\, S(\tau)\, d\tau \right\| \leqslant \int_t^{t+\sigma} \|\tilde{\mathscr{L}}(\tau)\|\, \|S(\tau)\|\, d\tau \leqslant$$

$$\leqslant \int_t^{t+\sigma} \|\tilde{\mathscr{L}}(\tau)\| d\tau \quad \varepsilon\beta^{1/2} \|\tilde{\xi}(t+\sigma)\|. \quad (29.191)$$

Puisque d'après (182), $\widetilde{\mathscr{L}}(\tau)$ est la matrice de Gram,

$$\int_t^{t+\sigma} \|\widetilde{\mathscr{L}}(\tau)\|\, d\tau = \left\| \int_t^{t+\sigma} \widetilde{\mathscr{L}}(\tau)\, d\tau \right\| = \|\mathscr{W}(t,\ t+\sigma)\| \leqslant \beta, \qquad (29.192)$$

et l'inégalité (191) devient

$$\left\| \int_t^{t+\sigma} \widetilde{\mathscr{L}}(\tau)\, S(\tau)\, d\tau \right\| \leqslant \varepsilon\beta^{3/2} \|\widetilde{\xi}(t+\sigma)\|. \qquad (29.193)$$

Il résulte de (193) que

$$\left\| \left\langle \widetilde{\xi}(t+\sigma),\ \int_t^{t+\sigma} \widetilde{\mathscr{L}}(\tau)\, S(\tau)\, d\tau \right\rangle \right\| \leqslant \varepsilon\beta^{3/2} \|\widetilde{\xi}(t+\sigma)\|^2. \qquad (29.194)$$

Avec (186) et (194) on a

$$\eta^2 \|\widetilde{\xi}(t+\sigma)\|^2 \geqslant (\alpha - 2\varepsilon\beta^{3/2}) \|\widetilde{\xi}(t+\sigma)\|^2 \qquad (29.195)$$

et, par suite,

$$\eta^2 \geqslant \alpha - 2\varepsilon\beta^{3/2}. \qquad (29.196)$$

Il est aisé de voir que ε^2 et η^2 ne peuvent pas s'annuler à la fois. En effet, si $\varepsilon = 0$, (196) implique que $\eta^2 = \alpha > 0$ et donc $\delta^2 \geqslant \alpha > > 0$. Et si $\eta^2 = 0$, la relation (196) devient

$$0 \geqslant \alpha - 2\varepsilon\beta^{3/2},$$

et, par conséquent,

$$2\varepsilon\beta^{3/2} \geqslant \alpha,$$

ou

$$\varepsilon \geqslant \frac{\alpha}{2\beta^{3/2}}.$$

Dans ce cas

$$\delta^2 = \varepsilon^2 \geqslant \frac{\alpha^2}{4\beta^3} > 0.$$

Ainsi nous avons obtenu pour δ^2 l'estimation

$$\delta^2 \geqslant k^2 > 0 \quad \text{où} \quad k^2 = \min\left(\alpha,\ \frac{\alpha^2}{4\beta^3}\right).$$

De la sorte, la relation (176) s'écrit

$$V(\widetilde{\xi}(t+\sigma),\ t+\sigma) - V(\widetilde{\xi}(t),\ t) \geqslant \geqslant k^2 \|\widetilde{\xi}(t+\sigma)\|^2 > 0, \qquad (29.197)$$

ce qui prouve la validité de la relation (155), c'est-à-dire la stabilité asymptotique uniforme du filtre optimal. Le théorème est démontré.

Théorème 6. *Soit le système* (74)

$$\frac{dx}{dt}=A(t)\,x+B(t)\,w(t),$$
$$z(t)=C(t)\,x(t)+v(t)$$

uniformément complètement observable et uniformément complètement commandable.

Alors, les matrices $\Sigma^{(a)}(t;\ \Sigma_0^{(a)},\ t_0)$ *et* $\Sigma^{(b)}(t;\ \Sigma_0^{(b)},\ t_0)$, *qui sont des solutions de l'équation de Riccati* (28.80) *sous des conditions initiales respectives* $\Sigma^{(a)}(t_0)=\Sigma_0^{(a)}$ *et* $\Sigma^{(b)}(t_0)=\Sigma_0^{(b)}$, *où* $\Sigma_0^{(a)}$ *et* $\Sigma_0^{(b)}$ *sont des matrices définies non négatives, vérifient la condition :*

$$\|\Sigma^{(a)}(t+\sigma;\ \Sigma_0^{(a)},\ t_0)-\Sigma^{(b)}(t+\sigma;\ \Sigma_0^{(b)},\ t_0)\|\leqslant$$
$$\leqslant c_1^2e^{-2c_2\sigma}\,\|\Sigma^{(a)}(t;\ \Sigma_0^{(a)},\ t_0)-\Sigma^{(b)}(t;\ \Sigma_0^{(b)},\ t_0\,\|$$
$$\text{pour tout } t\geqslant t_0. \qquad (29.198)$$

Ici c_1 *et* c_2 *sont des constantes positives qui font partie de l'expression* (155).

Démonstration. Pour condenser l'écriture, introduisons les notations

$$\Sigma_1=\Sigma^{(a)}(t;\ \Sigma_0^{(a)},\ t_0),\quad \Sigma_2=\Sigma^{(b)}(t;\ \Sigma_0^{(b)},\ t_0), \qquad (29.199)$$
$$\Delta(t)=\Sigma_1-\Sigma_2. \qquad (29.200)$$

Conformément à (28.80), on a

$$\frac{d\Sigma_1}{dt}=A(t)\,\Sigma_1+\Sigma_1A^*(t)-\Sigma_1C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma_1+B(t)\,Q(t)\,B^*(t),$$
$$\frac{d\Sigma_2}{dt}=A(t)\,\Sigma_2+\Sigma_2A^*(t)-\Sigma_2C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma_2+B(t)\,Q(t)\,B^*(t).$$
$$(29.201)$$

L'expression (201) conduit à

$$\frac{d}{dt}\Delta(t)=\mathcal{A}_1(t)\,\Delta(t)+\Delta(t)\,\mathcal{A}_2^*(t), \qquad (29.202)$$

avec

$$\mathcal{A}_i(t)=A(t)-\Sigma_iC^*(t)\,R^{-1}(t)\,C(t),\qquad i=1,\ 2. \qquad (29.203)$$

On peut vérifier directement que la solution de l'équation différentielle matricielle (202) est de la forme

$$\Delta(t)=\Psi_1(t,\ t_0)\,\Delta(t_0)\,\Psi_2^*(t,\ t_0), \qquad (29.204)$$

où, d'après (151),

$$\Psi_i(t, t_0) = \varkappa_i(t)\, \varkappa_i^{-1}(t_0), \qquad i = 1, 2 \tag{29.205}$$

et où $\varkappa_i(t)$ est la matrice fondamentale des solutions de l'équation différentielle vectorielle

$$\frac{d\rho_i}{dt} = \mathcal{A}_i(t)\, \rho_i, \qquad i = 1, 2, \tag{29.206}$$

c'est-à-dire $\varkappa_1(t)$ et $\varkappa_2(t)$ sont des matrices fondamentales des solutions des équations qui constituent l'équation du filtre optimal (149) pour $\Sigma(t) = \Sigma^{(a)}(t; \Sigma_0^{(a)}, t_0)$ et $\Sigma(t) = \Sigma^{(b)}(t; \Sigma_0^{(b)}, t_0)$ respectivement.

Il résulte de (204) que

$$\Delta(t + \sigma) = \Psi_1(t + \sigma, t)\, \Delta(t)\, \Psi_2^*(t + \sigma, t). \tag{29.207}$$

D'où, en vertu de (155),

$$\| \Delta(t + \sigma) \| \leqslant c_1^2 e^{-2c_2\sigma} \| \Delta(t) \|, \tag{29.208}$$

ce qu'il fallait démontrer.

T h é o r è m e 7. *Soit le système* (74) *uniformément complètement observable et uniformément complètement commandable. La solution de l'équation de Riccati* (28.80), *associée à la condition initiale* $\Sigma(t_0) = 0$, *possède alors pour* $t_0 \to -\infty$ *la limite*

$$\lim_{t_0 \to -\infty} \Sigma(t; 0, t_0) = \widetilde{\Sigma}(t) \tag{29.209}$$

qui existe pour tout t.

D é m o n s t r a t i o n. Considérons deux solutions de l'équation de Riccati (28.80) qui correspondent à des conditions initiales différentes

$$\Sigma^{(a)}(t; 0, t_0) = \Sigma(t; 0, t_0) \tag{29.210}$$

et

$$\Sigma^{(b)}(t; 0, t_1) = \Sigma(t; 0, t_1), \tag{29.211}$$

où

$$t_0 < t_1 < t. \tag{29.212}$$

Si l'on désigne

$$\Sigma^{(a)}(t_1) = \Sigma^{(a)}(t_1; 0, t_0),$$

la solution (210) peut s'écrire

$$\Sigma^{(a)}(t; 0, t_0) = \Sigma^{(a)}(t; \Sigma^{(a)}(t_1), t_1). \tag{29.213}$$

En vertu du lemme 2, pour $t_1 \geqslant t_0 + \sigma$, la matrice $\Sigma^{(a)}(t_1)$ est définie positive. D'après (211), la matrice $\Sigma^{(b)}(t_1) = 0$.

En raison du théorème 6 et conformément à (207), on a

$$\Sigma^{(a)}(t; \Sigma^{(a)}(t_1), t_1) - \Sigma^{(b)}(t; 0, t_1) = \\ = \Psi_1(t, t_1)\, \Sigma^{(a)}(t_1)\, \Psi_2^*(t, t_1). \tag{29.214}$$

Le second membre de l'expression (214) est une matrice définie positive ; donc

$$\Sigma^{(a)}(t\,;\ \Sigma^{(a)}(t_1),\ t_1) > \Sigma^{(b)}(t\,;\ 0,\ t_1),$$

ou, d'après (210), (211) et (213),

$$\Sigma\ (t;\ 0,\ t_0) > \Sigma\ (t;\ 0,\ t_1), \qquad t_0 < t_1 < t. \qquad (29.215)$$

Ainsi, la matrice $\Sigma\ (t;\ 0,\ t_0)$ ne décroît pas avec la diminution de t_0, et d'après (142), elle est bornée supérieurement. On en tire, en vertu du théorème de la convergence monotone d'une série des opérateurs positifs dans un espace de Hilbert, que pour $t_0 \to -\infty$ la matrice $\Sigma\ (t;\ 0,\ t_0)$ tend vers la limite

$$\lim_{t_0 \to -\infty} \Sigma\,(t\,;\ \ 0,\ t_0) = \tilde{\Sigma}\,(t),$$

qui existe pour tout t. Le théorème est démontré.

Théorème 8. *Soit le système* (74) *uniformément complètement observable et uniformément complètement commandable. Toute solution* $\Sigma\ (t;\ \Sigma_0,\ t_0)$ *de l'équation de Riccati* (28.80) *vérifiant la condition initiale* $\Sigma\ (t_0) = \Sigma_0$, *où* $t_0 > -\infty$ *et où* Σ_0 *est une matrice définie non négative, converge alors uniformément vers* $\tilde{\Sigma}\ (t)$, *où la matrice* $\tilde{\Sigma}\ (t)$ *est déterminée par l'expression* (209). *En ce sens* $\tilde{\Sigma}\ (t)$ *est un état d'équilibre mobile qui vérifie* (28.80).

Démonstration. Considérons l'équation différentielle homogène (149) obtenue de l'équation différentielle du filtre optimal pour $z\ (t) \equiv 0$:

$$\frac{d\rho}{dt} = [A\,(t) - \Sigma\,(t)\,C^*\,(t)\,R^{-1}\,(t)\,C\,(t)]\,\rho.$$

D'après ce quiaé té démontré précédemment, la solution triviale de (149) est uniformément asymptotiquement stable.

Considérons la fonction

$$V\ (\rho,\ t)\ =\ \rho^*\Sigma^{-1}\ (t)\ \rho. \qquad (29.216)$$

Sa dérivée par rapport au temps $\frac{d\tilde{V}}{dt}$ est en vertu de l'équation différ·ntielle (149) et, conformément à (143),

$$\frac{d\tilde{V}}{dt} = -\rho^*\,[\Sigma^{-1}\,(t)\,B\,(t)\,Q\,(t)\,B^*\,(t)\,\Sigma^{-1}\,(t) + \\ + C^*\,(t)\,R^{-1}\,(t)\,C\,(t)]\,\rho. \qquad (29.217)$$

La matrice entre crochets du second membre de (217) étant une matrice de Gram, la forme quadratique du second membre de (217) ne peut être positive et la croissance de la fonction $\tilde{V}\ (\rho,\ t)$ avec

l'augmentation de t est impossible. Ainsi,

$$\widetilde{V}(\rho(t+\sigma),\ t+\sigma)-\widetilde{V}(\rho(t),\ t)\leqslant 0 \qquad (29.218)$$

ou

$$\rho^*(t+\sigma)\,\Sigma^{-1}(t+\sigma)\,\rho(t+\sigma)\leqslant\rho^*(t)\,\Sigma^{-1}(t)\,\rho(t). \qquad (29.219)$$

La solution triviale de l'équation (149) étant uniformément asymptotiquement stable,

$$\|\rho(t+\sigma)\|<\|\rho(t)\|. \qquad (29.220)$$

On déduit de (219) et (220) que

$$\Sigma^{-1}(t+\sigma)\leqslant\Sigma^{-1}(t), \qquad (29.221)$$

d'où, conformément à (146) et (147),

$$\Sigma(t+\sigma)\geqslant\Sigma(t).$$

De cette façon, la matrice $\Sigma(t;\Sigma_0, t_0)$ ne décroît pas avec l'augmentation de t. De plus, d'après (142), la matrice $\Sigma(t;\ \Sigma_0,\ t_0)$ est bornée supérieurement. On en tire que toute solution $\Sigma(t;\ \Sigma_0,\ t_0)$ de l'équation de Riccati (28.80), où Σ_0 est une matrice définie non négative, converge uniformément vers $\widetilde{\Sigma}(t)$, où $\widetilde{\Sigma}(t)$ est la matrice définie par l'expression (209). Le théorème est démontré.

Systèmes stationnaires. Dans le cas où (74) est un système stationnaire, c'est-à-dire où A, B, C, Q et R sont des matrices constantes, la solution de l'équation de Riccati (28.80) ne dépend que de la différence $t-t_0$. Dans ce cas, $\widetilde{\Sigma}(t)=\hat{\Sigma}=$ $=$ const.

Notons encore que les systèmes complètement commandables (observables) *stationnaires* sont en même temps des systèmes uniformément complètement commandables (observables).

Théorème 9. *Supposons que le système* (74) *soit stationnaire, c'est-à-dire que A, B et C sont des matrices constantes, alors que les matrices Q et R, des matrices constantes définies positives. Supposons également que le système stationnaire considéré* (74) *est uniformément complètement commandable et uniformément complètement observable, c'est-à-dire qu'il vérifie les conditions*

$$\begin{aligned}\text{rang}\,[B\quad AB\ \ldots\ A^{n-1}B]=n,\\ \text{rang}\,[C^*\quad A^*C^*\ \ldots\ A^{*n-1}C^*]=n.\end{aligned} \qquad (29.222)$$

Alors, les résultats suivants ont lieu:

1°. *La solution $\Sigma(t;\ 0,\ t_0)$ de l'équation de Riccati*

$$\frac{d\Sigma}{dt}=A\Sigma+\Sigma A^*-\Sigma C^*R^{-1}C\Sigma+BQB^* \qquad (29.223)$$

possède pour $t_0 \to -\infty$ la limite

$$\lim_{t_0 \to -\infty} \Sigma(t;\ 0,\ t_0) = \hat{\Sigma} \tag{29.224}$$

qui existe pour tout t; de plus, $\hat{\Sigma}$ est une matrice constante symétrique définie positive qui est solution de l'équation algébrique matricielle non linéaire

$$A\Sigma + \Sigma A^* - \Sigma C^* R^{-1} C\Sigma + BQB^* = 0. \tag{29.225}$$

2°. *Toute solution $\Sigma(t;\ 0,\ t_0)$ de l'équation de Riccati* (223) *vérifiant la condition initiale $\Sigma(t_0) = \Sigma_0$, où $t_0 > -\infty$, et où Σ_0 est une matrice définie non négative, tend avec $t \to \infty$ uniformément vers $\hat{\Sigma}$; c'est-à-dire la matrice $\hat{\Sigma}$ est l'état d'équilibre défini positif unique vérifiant l'équation différentielle de Riccati* (223).

Démonstration. La première partie du théorème, c'est-à-dire la proposition (224), se déduit du théorème 7. La deuxième partie se déduit du théorème 8. La matrice $\hat{\Sigma}$ est définie positive d'après le lemme 2. L'unicité de la solution constante de l'équation (223) se déduit du fait que plus d'une solution constante contredirait le théorème 6. Le théorème est démontré.

3. Algorithme de construction de la solution limite de l'équation de Riccati pour les systèmes stationnaires à temps infini d'observation. Considérons l'équation de Riccati (28.80)

$$\frac{d\Sigma}{dt} = A(t)\,\Sigma + \Sigma A^*(t) - \Sigma C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma + B(t)\,Q(t)\,B^*(t) \qquad \Sigma(t_0) = \Sigma_0 \geqslant 0.$$

Pour les systèmes stationnaires, les matrices A, B, C, Q, R sont constantes et l'équation (28.80) s'écrit

$$\frac{d\Sigma}{dt} = A\Sigma + \Sigma A^* - \Sigma C^* R^{-1} C\Sigma + BQB^*, \qquad \Sigma(t_0) = \Sigma_0 > 0. \tag{29.226}$$

L'équation du filtre optimal (28.82) d'un système stationnaire devient

$$\frac{d\rho}{dt} = (A - \Sigma C^* R^{-1} C)\,\rho + \Sigma C^* R^{-1} z(t), \quad \rho(t_0) = \bar{x}_0. \tag{29.227}$$

Introduisons les notations

$$\mathfrak{R} = C^* R^{-1} C, \tag{29.228}$$

$$L = BQB^*.$$

L'équation de Riccati (226) se met alors sous la forme

$$\frac{d\Sigma}{dt} = A\Sigma + \Sigma A^* - \Sigma\mathfrak{R}\Sigma + L, \quad \Sigma(t_0) = \Sigma_0 > 0. \tag{29.229}$$

L'équation du filtre optimal (227) devient

$$\frac{d\rho}{dt} = (A - \Sigma\mathfrak{R})\,\rho + \Sigma C^* R^{-1} z(t), \quad \rho(t_0) = \bar{x}_0. \tag{29.230}$$

Le système d'équations différentielles vectorielles (28.85) du système stationnaire concerné est

$$\frac{d\xi}{dt} = -A^*\xi + \mathfrak{N}\mu,$$
$$\frac{d\mu}{dt} = L\xi + A\mu, \qquad (29.231)$$

où $\xi(t)$ et $\mu(t)$ sont les vecteurs de dimension n. 101

Désignons par $\mathfrak{M}$ la matrice partitionnée

$$\mathfrak{M} = \begin{bmatrix} -A^* & \mathfrak{N} \\ L & A \end{bmatrix}. \qquad (29.232)$$

Par $S(\Sigma)$ on désigne le second membre de l'équation différentielle matricielle de Riccati (229)

$$S(\Sigma) = A\Sigma + \Sigma A^* - \Sigma\mathfrak{N}\Sigma + L. \qquad (29.233)$$

Comme nous l'avons montré plus haut, si le système initial (74) est stationnaire, s'il vérifie les conditions d'observabilité et de commandabilité (222), Q et R étant des matrices constantes définies positives, alors pour $t \to \infty$ la solution limite de l'équation (229) est la matrice $\hat{\Sigma}$ symétrique définie positive, solution de l'équation algébrique matricielle non linéaire

$$S(\Sigma) = 0. \qquad (29.234)$$

La solution $\Sigma = \hat{\Sigma}$ (où $\hat{\Sigma}$ est une matrice symétrique définie positive) de (234) existe et est unique.

L'un des algorithmes possibles de construction de la matrice $\hat{\Sigma}$ est donné par le théorème qui suit [19].

Théorème. *La matrice symétrique définie positive $\hat{\Sigma}$, solution de l'équation algébrique matricielle* (234)

$$S(\Sigma) = 0,$$

est déterminée par la relation

$$[-\hat{\Sigma},\ E]\,\Delta\mathfrak{M} = 0 \qquad (29.235)$$

ou par la relation équivalente

$$\Delta(-\mathfrak{M})\begin{bmatrix} E \\ \hat{\Sigma} \end{bmatrix} = 0, \qquad (29.236)$$

avec E, la matrice unité. Ici, $\Delta(\mathfrak{M})$ est le polynôme dont la variable est la matrice $\mathfrak{M}$. Le polynôme scalaire correspondant $\Delta(\lambda)$ est déterminé par la relation

$$\det(\lambda E - \mathfrak{M}) = (-1)^n\,\Delta(\lambda)\,\Delta(-\lambda), \qquad (29.237)$$

le polynôme $\Delta(\lambda)$ étant de degré n et ses zéros possédant des parties réelles négatives.

Démonstration. Comme nous l'avons montré précédemment, la matrice symétrique définie positive $\hat{\Sigma}$ qui vérifie l'équation algébrique matricielle (234) existe et est unique.

D'après (230), à la matrice $\Sigma = \hat{\Sigma}$ correspond le filtre optimal décrit par l'équation différentielle

$$\frac{d\rho}{dt} = (A - \hat{\Sigma}\mathfrak{N})\,\rho + \hat{\Sigma}C^*R^{-1}z(t), \quad \rho(t_0) = \bar{x}_0. \qquad (29.238)$$

Le filtre optimal (238) est asymptotiquement stable, c'est-à-dire la matrice

$$\bar{A} = A - \hat{\Sigma}\mathfrak{N} \qquad (29.239)$$

est asymptotiquement stable. Désignons par $\Delta(\lambda)$ le polynôme caractéristique de la matrice $\bar{A}$

$$\Delta(\lambda) = \det(\lambda E - \bar{A}). \tag{29.240}$$

$\bar{A}$ étant une matrice asymptotiquement stable, tous les zéros du polynôme $\Delta(\lambda)$ possèdent des parties réelles négatives.

Désignons par T la matrice

$$T = \begin{bmatrix} 0 & E \\ E & \hat{\Sigma} \end{bmatrix}. \tag{29.241}$$

Son inverse T^{-1} est de la forme

$$T^{-1} = \begin{bmatrix} -\hat{\Sigma} & E \\ E & 0 \end{bmatrix}. \tag{29.242}$$

En outre, on a

$$T^{-1}\mathfrak{M}T = \begin{bmatrix} -\hat{\Sigma} & E \\ E & 0 \end{bmatrix} \begin{bmatrix} -A^* & \mathfrak{N} \\ L & A \end{bmatrix} \begin{bmatrix} 0 & E \\ E & \hat{\Sigma} \end{bmatrix} = \begin{bmatrix} A-\hat{\Sigma}\mathfrak{N} & S(\hat{\Sigma}) \\ \mathfrak{N} & -(A-\hat{\Sigma}\mathfrak{N})^* \end{bmatrix}. \tag{29.243}$$

Puisque

$$S(\hat{\Sigma}) = 0,$$

conformément à (239), la matrice (243) devient

$$T^{-1}\mathfrak{M}T = \begin{bmatrix} \bar{A} & 0 \\ \mathfrak{N} & -\bar{A}^* \end{bmatrix}. \tag{29.244}$$

Comme

$$\det(\lambda E - \mathfrak{M}) = \det(\lambda E - T^{-1}\mathfrak{M}T), \tag{29.245}$$

on obtient

$$\det(\lambda E - \mathfrak{M}) = \det \begin{bmatrix} \lambda E - \bar{A} & 0 \\ -\mathfrak{N} & \lambda E + \bar{A}^* \end{bmatrix} = \det(\lambda E - \bar{A}) \det(\lambda E + \bar{A}^*). \tag{29.246}$$

Plus haut nous avons noté (240)

$$\det(\lambda E - \bar{A}) = \Delta(\lambda).$$

Puisque $\det(\lambda E + \bar{A}^*)$ est le déterminant caractéristique du système

$$\frac{d\tilde{\xi}}{dt} = -(A - \hat{\Sigma}\mathfrak{N})^* \tilde{\xi}, \tag{29.247}$$

conjugué du filtre optimal (238), il vient

$$\det(\lambda E + \bar{A}^*) = (-1)^n \Delta(-\lambda). \tag{29.248}$$

Nous avons ainsi démontré la proposition (237) du théorème suivant laquelle

$$\det(\lambda E - \mathfrak{M}) = (-1)^n \Delta(\lambda) \Delta(-\lambda)$$

et établi que $\Delta\lambda$ est déterminé par la relation (240).

Ecrivons la matrice $\Delta\mathfrak{M}$, composée en remplaçant dans le polynôme $\Delta(\lambda)$ la variable scalaire λ par la matrice carrée $\mathfrak{M}$, sous la forme partitionnée

$$\Delta(\mathfrak{M}) = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{Z} \end{bmatrix}, \tag{29.249}$$

où $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ sont des matrices $n \times n$. De plus, la matrice $T^{-1}\Delta(\mathfrak{M})T$ est de la forme

$$T^{-1}\Delta(\mathfrak{M})T = \begin{bmatrix} -\hat{\Sigma} & E \\ E & 0 \end{bmatrix} \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \begin{bmatrix} 0 & E \\ E & \hat{\Sigma} \end{bmatrix} =$$

$$= \begin{bmatrix} \mathcal{D} - \hat{\Sigma}\mathcal{B}, & -\hat{\Sigma}\mathcal{A} + \mathcal{C} + (-\hat{\Sigma}\mathcal{B} + \mathcal{D})\Sigma \\ \mathcal{B} & \mathcal{A} + \mathcal{B}\hat{\Sigma} \end{bmatrix}. \quad (29.250)$$

Puisque d'après (244)

$$T^{-1}\mathfrak{M}T = \begin{bmatrix} \bar{A} & 0 \\ \mathfrak{N} & -\bar{A}^* \end{bmatrix}$$

et comme $\Delta(\lambda)$ est le polynôme de λ, on a

$$\Delta(T^{-1}\mathfrak{M}T) = \Delta\left(\begin{bmatrix} \bar{A} & 0 \\ \mathfrak{N} & -\bar{A}^* \end{bmatrix}\right) = \begin{bmatrix} \Delta(\bar{A}) & 0 \\ a & \Delta(-\bar{A}^*) \end{bmatrix}, \quad (29.251)$$

où a est une matrice $n \times n$, en général non nulle.

Etant donné que d'après (240)

$$\Delta(\lambda) = \det(\lambda E - \bar{A}),$$

en vertu du théorème de Hamilton-Cayley

$$\Delta(\bar{A}) = 0, \quad (29.252)$$

et ainsi la matrice (251) devient

$$\Delta(T^{-1}\mathfrak{M}T) = \begin{bmatrix} 0 & 0 \\ a & \Delta(-\bar{A}^*) \end{bmatrix}. \quad (29.253)$$

Constatons que d'après (248)

$$\det(\lambda E + \bar{A}^*) = (-1)^n \, \Delta(-\lambda);$$

c'est pourquoi le théorème d'Hamilton-Cayley conduit à

$$\Delta(-(-\bar{A}^*)) = \Delta(\bar{A}^*) = 0. \quad (29.254)$$

Comme le filtre optimal est asymptotiquement stable, les zéros du polynôme $\Delta(\lambda)$ reposent dans le plan de la variable complexe λ strictement à gauche de l'axe imaginaire, c'est-à-dire $\Delta(\lambda)$ est un polynôme de Hurwitz. C'est pourquoi dans le polynôme $\Delta(\lambda)$ les coefficients affectés aux puissances paires de même qu'aux puissances impaires de λ ne peuvent être nuls et on a d'après la relation (254)

$$\Delta(-\bar{A}^*) \neq 0. \quad (29.255)$$

Déduisons maintenant une relation auxiliaire. Conformément à (240), le polynôme scalaire $\Delta(\lambda)$ est de la forme

$$\Delta(\lambda) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n. \quad (29.256)$$

Il en résulte que

$$\Delta(\mathfrak{M}) = \mathfrak{M}^n + a_1\mathfrak{M}^{n-1} + \ldots + a_{n-1}\mathfrak{M} + a_nE. \quad (29.257)$$

Ici E est la matrice unité, et les coefficients $a_1, \ldots, a_{n-1}, a_n$ sont des scalaires.

D'après (257), la matrice $T^{-1}\Delta(\mathfrak{M})T$ s'écrit

$$T^{-1}\Delta(\mathfrak{M})T = T^{-1}\mathfrak{M}^nT + a_1T^{-1}\mathfrak{M}^{n-1}T + \ldots + a_{n-1}T^{-1}\mathfrak{M}T + a_nE. \quad (29.258)$$

Comme

$$(T^{-1}\mathfrak{M}T)^2 = T^{-1}\mathfrak{M}TT^{-1}\mathfrak{M}T = T^{-1}\mathfrak{M}^2T,$$
$$(T^{-1}\mathfrak{M}T)^3 = T^{-1}\mathfrak{M}TT^{-1}\mathfrak{M}TT^{-1}\mathfrak{M}T = T^{-1}\mathfrak{M}^3T,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$(T^{-1}\mathfrak{M}T)^s = T^{-1}\mathfrak{M}^sT, \tag{29.259}$$

la matrice $\Delta\,(T^{-1}\mathfrak{M}T)$ est, en vertu de (257),

$$\Delta\,(T^{-1}\mathfrak{M}T) = (T^{-1}\mathfrak{M}T)^n + a_1\,(T^{-1}\mathfrak{M}T)^{n-1} + \ldots + a_{n-1}T^{-1}\mathfrak{M}T + a_nE = T^{-1}\mathfrak{M}^nT + a_1T^{-1}\mathfrak{M}^{n-1}T + \ldots + a_{n-1}T^{-1}\mathfrak{M}T + a_nE. \tag{29.260}$$

Les relations (258) et (260) entraînent que

$$T^{-1}\Delta\,(\mathfrak{M})\,T = \Delta\,(T^{-1}\mathfrak{M}T). \tag{29.261}$$

C'est précisément la relation auxiliaire qui nous intéresse.

On tire de (261), (250) et (253) que

$$\mathscr{D} = \hat{\Sigma}\mathscr{B}, \quad \mathscr{C} = \hat{\Sigma}\mathscr{A}, \tag{29.262}$$

d'où

$$\hat{\Sigma} = \mathscr{D}\mathscr{B}^{-1} = \mathscr{C}\mathscr{A}^{-1}. \tag{29.263}$$

Les relations (262) sont équivalentes à la relation (235) donnée par le théorème. En effet, d'après (235),

$$[-\hat{\Sigma},\ E]\ \Delta\,(\mathfrak{M}) = 0,$$

ou, d'après (249),

$$[-\hat{\Sigma},\ E]\begin{bmatrix}\mathscr{A} & \mathscr{B}\\ \mathscr{C} & \mathscr{D}\end{bmatrix} = [-\hat{\Sigma}\mathscr{A}+\mathscr{C},\ -\hat{\Sigma}\mathscr{B}+\mathscr{D}] = 0,$$

ce qui coïncide précisément avec les relations (262).

Pour justifier la relation (236) du théorème, on procède d'une façon analogue. Le théorème est démontré.

§ 30. Filtrage optimal des bruits corrélés

1. Systèmes à perturbations intercorrélées. Considérons le problème du filtrage optimal d'un système linéaire non stationnaire

$$\frac{dx}{dt} = A\,(t)\,x + B\,(t)\,w\,(t),$$
$$z\,(t) = C\,(t)\,x\,(t) + v\,(t). \tag{30.1}$$

Ici $w\,(t)$ et $v\,(t)$ sont des bruits blancs gaussiens à moyennes nulles

$$\mathrm{M}\,[w\,(t)] = 0,\ \mathrm{M}\,[v\,(t)] = 0 \text{ pour tout } t. \tag{30.2}$$

A la différence des hypothèses faites dans les paragraphes précédents, nous allons considérer que les processus aléatoires $w\,(t)$

et $v(t)$ sont intercorrélés et que leurs fonctions de corrélation sont

$$\begin{aligned} \operatorname{cov}[w(t);\ w(\tau)] &= \mathbf{M}[w(t)\, w^*(\tau)] = \delta(t-\tau)\, Q(t), \\ \operatorname{cov}[v(t);\ v(\tau)] &= \mathbf{M}[v(t)\, v^*(\tau)] = \delta(t-\tau)\, R(t), \\ \operatorname{cov}[w(t);\ v(\tau)] &= \mathbf{M}[w(t)\, v^*(\tau)] = \delta(t-\tau)\, \mathscr{L}(t), \end{aligned} \tag{30.3}$$

où $Q(t)$ est la matrice symétrique définie non négative, et $R(t)$ la matrice symétrique définie positive.

Soit t_0 l'instant initial ($t_0 > -\infty$), et le vecteur $x(t_0)$, l'état initial du système. Supposons que $x(t_0)$ est une variable aléatoire vectorielle gaussienne non corrélée avec $w(t)$ et $v(t)$ à moyenne

$$\mathbf{M}[x(t_0)] = 0. \tag{30.4}$$

Sa matrice de corrélation

$$\operatorname{cov}[x(t_0);\ x(t_0)] = \mathbf{M}[x(t_0)\, x^*(t_0)] = \Sigma_0 \tag{30.5}$$

est une matrice définie non négative supposée connue.

A la différence de (28.82) et (28.80), dans ce problème, l'équation différentielle du filtre optimal s'écrit

$$\frac{d\rho}{dt} = A(t)\,\rho + \widetilde{G}(t)\,[z(t) - C(t)\,\rho], \quad \rho(t_0) = 0. \tag{30.6}$$

Ici

$$\widetilde{G}(t) = [\Sigma(t)\, C^*(t) + B(t)\, \mathscr{L}(t)]\, R^{-1}(t) \tag{30.7}$$

et $\Sigma(t)$ est la solution de l'équation différentielle matricielle de Riccati

$$\begin{aligned} \frac{d\Sigma}{dt} = A(t)\,\Sigma + \Sigma A^*(t) - [\Sigma C^*(t) + B(t)\,\mathscr{L}(t)] \times \\ \times R^{-1}(t)\,[C(t)\,\Sigma + \mathscr{L}^*(t)\, B^*(t)] + \\ + B(t)\, Q(t)\, B^*(t), \quad \Sigma(t_0) = \Sigma_0. \end{aligned} \tag{30.8}$$

Montrons que l'équation de Riccati (8) peut être ramenée à la forme analogue à celle de (28.80). A cet effet, introduisons les notations suivantes

$$\begin{aligned} \hat{\mathscr{A}}(t) &= A(t) - B(t)\, \mathscr{L}(t)\, R^{-1}(t)\, C(t), \\ \hat{Q}(t) &= Q(t) - \mathscr{L}(t)\, R^{-1}(t)\, \mathscr{L}^*(t). \end{aligned} \tag{30.9}$$

Supposons ensuite que $\hat{Q}(t)$ est une matrice définie non négative. En portant dans l'équation (8) les valeurs de $A(t)$ et $Q(t)$, données par (9), on ramène l'équation (8) à la forme

$$\begin{aligned} \frac{d\Sigma}{dt} = \hat{\mathscr{A}}(t)\,\Sigma + \Sigma \hat{\mathscr{A}}^*(t) - \Sigma C^*(t)\, R^{-1}(t)\, C(t)\,\Sigma + \\ + B(t)\, \hat{Q}(t)\, B^*(t), \quad \Sigma(t_0) = \Sigma_0. \end{aligned} \tag{30.10}$$

Conformément à (9), l'équation différentielle (6) du filtre optimal devient

$$\frac{d\rho}{dt} = [\hat{\mathscr{A}}(t) - \Sigma(t) C^*(t) R^{-1}(t) C(t)]\rho +$$
$$+ [\Sigma(t) C^*(t) R^{-1}(t) + B(t) \mathscr{L}(t) R^{-1}(t)] z(t), \quad \rho(t_0) = 0. \quad (30.11)$$

L'équation différentielle de Riccati (10) ne se distingue de (28.80) que par les paramètres $A(t)$ et $Q(t)$ remplacés respectivement par $\hat{\mathscr{A}}(t)$ et $\hat{Q}(t)$. La correspondance est la même pour les équations différentielles homogènes formées à partir de (28.82) et (11) pour $z(t) \equiv 0$. C'est pourquoi les résultats obtenus au § 29 sur les propriétés des solutions de l'équation de Riccati (28.80) et sur la stabilité du filtre optimal concernent également notre problème. Certes, les fonctions $A(t)$ et $Q(t)$ doivent être remplacées par les fonctions respectives $\hat{\mathscr{A}}(t)$ et $\hat{Q}(t)$.

2. Systèmes à perturbations de mesure autres que les bruits blancs. Dans les sections précédentes, les filtres optimaux de Kalman-Bucy ont été construits pour le cas où les perturbations étaient des bruits blancs. La généralisation au cas où les mesures du signal de sortie sont entachées de bruits dus au passage du bruit blanc par un système linéaire non stationnaire a été étudiée par Bucy [19].

Examinons le système décrit par les équations vectorielles

$$\begin{aligned} &\frac{dx}{dt} = A(t)x + B(t)w(t), && x(t_0) = c, \\ &\frac{dv}{dt} = S(t)v + \varkappa(t), && v(t_0) = b, \qquad (30.12) \\ &z(t) = Cx(t) + v(t). \end{aligned}$$

Ici x est le vecteur d'état de dimension n ; $w(t)$, le vecteur de dimension r à l'entrée du système ; $A(t)$, la matrice $n \times n$; $B(t)$, la matrice $n \times r$; $\varkappa(t)$, le vecteur de dimension m ; v, le vecteur de dimension m ; $S(t)$, la matrice $m \times m$; $z(t)$, le vecteur de dimension m ; C, la matrice constante $m \times n$.

L'entrée $w(t)$ est un bruit blanc gaussien à moyenne nulle

$$\mathrm{M}[w(t)] = 0 \qquad (30.13)$$

et à matrice de corrélation

$$\operatorname{cov}[w(t; w(\tau)] = \mathrm{M}[w(t) w^*(\tau)] = \delta(t - \tau) Q(t), \qquad (30.14)$$

où $\delta(t)$ est la fonction impulsion de Dirac et $Q(t)$, la matrice symétrique définie non négative $r \times r$.

Le signal $\varkappa(t)$ est un bruit blanc gaussien à moyenne nulle

$$\mathrm{M}[\varkappa(t)] = 0 \qquad (30.15)$$

et à matrice de corrélation

$$\operatorname{cov}[\varkappa(t);\ \varkappa(\tau)] = \mathrm{M}[\varkappa(t)\,\varkappa^*(\tau)] = \delta(t-\tau)\,R(t), \quad (30.16)$$

où $R(t)$ est la matrice symétrique définie positive $m \times m$.

Le vecteur c est une grandeur aléatoire vectorielle gaussienne de dimension n indépendante de $w(t)$, à moyenne nulle et à matrice de corrélation connue

$$\mathrm{M}[cc^*] = \Gamma. \quad (30.17)$$

Le vecteur b est une grandeur aléatoire vectorielle gaussienne de dimension m indépendante de $\varkappa(t)$, à moyenne nulle et à matrice de corrélation connue

$$\mathrm{M}[bb^*] = \mathscr{K}. \quad (30.18)$$

On suppose que $w(t)$, $\varkappa(t)$, c et b ne soient pas corrélés.

Désignons par $\hat{x}(t)$ l'estimation optimale de l'état $x(t)$ du système (12), c'est-à-dire celle qui minimise la fonctionnelle (28.24), d'après le vecteur fonction $z(t)$ accessible à la mesure dans l'intervalle de temps $[t_0, T]$.

D'une façon analogue à (26.116), l'estimation optimale $\hat{x}(t)$ vérifie le lemme sur la projection orthogonale dans un espace de Hilbert, qui dans ce problème conduit à la relation

$$\mathrm{M}\{[x(t) - \hat{x}(t)]\,z^*(\tau)\} = 0, \qquad t_0 \leqslant \tau < t. \quad (30.19)$$

Désignons par $\mathscr{D}(t)$ et $\mathscr{U}$ les matrices symétriques $m \times m$

$$\mathscr{D}(t) = CB(t)\,Q(t)\,B^*(t)\,C^* + R(t), \quad (30.20)$$

$$\mathscr{U} = C\Gamma C^* + \mathscr{K}. \quad (30.21)$$

Supposons dans ce qui suit que $\mathscr{D}(t)$ et $\mathscr{U}$ sont des matrices définies positives.

Comme l'a montré Bucy [19], l'estimation optimale $\hat{x}(t)$ de l'état $x(t)$ du système (12) s'écrit

$$\hat{x}(t) = \rho(t), \quad (30.22)$$

où $\rho(t)$ est la solution de l'équation différentielle

$$\frac{d\rho}{dt} = [A(t) - \mathscr{V}(t)\,CA(t) + \mathscr{V}(t)\,S(t)\,C]\,\rho + \\ + \mathscr{V}(t)\,\frac{dz}{dt} + \mathscr{V}(t)\,S(t)\,z(t), \quad (30.23)$$

$$\rho(t_0) = \Gamma C^* \mathscr{U}^{-1} z(t_0).$$

Nous avons désigné ici par $\mathscr{V}(t)$ la matrice $n \times m$

$$\mathscr{V}(t) = \{\mathscr{E}(t)\,[A^*(t)\,C^* - C^*S(t)] + B(t)\,Q(t)\,B^*(t)\,C^*\}\,\mathscr{D}^{-1}(t), \quad (30.24)$$

et par $\mathscr{E}(t)$, la solution de l'équation différentielle matricielle de Riccati

$$\frac{d\mathscr{E}}{dt} = A(t)\mathscr{E} + \mathscr{E}A^*(t) - \mathscr{V}(t)\mathscr{D}(t)\mathscr{V}^*(t) + B(t)Q(t)B^*(t), \quad (30.25)$$

qui vérifie la condition initiale

$$\mathscr{E}(t_0) = \Gamma - \Gamma C^*\mathscr{U}^{-1}C\Gamma = \mathscr{T}. \quad (30.26)$$

Dans ce problème, l'équation différentielle (23) est celle du *filtre optimal.*

Introduisons la notation $\gamma(t)$ pour le vecteur fonction de dimension n

$$\gamma(t) = \rho(t) - \mathscr{V}(t)z(t). \quad (30.27)$$

Désignons par $\mathscr{F}(t)$ et $\mathscr{N}(t)$ les matrices $m \times m$ et $m \times n$

$$\mathscr{F}(t) = A(t) - \mathscr{V}(t)CA(t) + \mathscr{V}(t)S(t)C, \quad (30.28)$$

$$\mathscr{N}(t) = \mathscr{F}(t)\mathscr{V}(t) - \mathscr{V}(t)S(t) - \frac{d\mathscr{V}}{dt}. \quad (30.29)$$

L'équation (23) du filtre optimal peut se mettre sous la forme

$$\begin{aligned} \frac{d\gamma}{dt} &= \mathscr{F}(t)\gamma + \mathscr{N}(t)z(t), \\ \gamma(t_0) &= [\Gamma C^*\mathscr{U}^{-1} - \mathscr{V}(t_0)]\,z(t_0). \end{aligned} \quad (30.30)$$

A la différence de (23), le second membre de l'équation (30) ne compte pas de dérivée par rapport au vecteur fonction $z(t)$ défini par observation.

Après avoir trouvé de l'équation différentielle (30) le vecteur fonction $\gamma(t)$, déterminons d'après (27) et (22) l'estimation à obtenir $\hat{x}(t)$ de l'état du système

$$\hat{x}(t) = \gamma(t) + \mathscr{V}(t)z(t). \quad (30.31)$$

Montrons encore que l'équation de Riccati (25) peut se ramener à une forme analogue à celle de l'équation (28.80). A cet effet, introduisons les notations

$$\begin{aligned} &\tilde{A}(t) = A(t) - B(t)Q(t)B^*(t)C^*\mathscr{D}^{-1}(t)\tilde{C}(t), \\ &\tilde{B}(t) = B(t), \quad \tilde{C}(t) = CA(t) - S(t)C, \\ &\tilde{Q}(t) = Q(t) - Q(t)B^*(t)C^*\mathscr{D}^{-1}(t)CB(t)Q(t), \\ &\tilde{R}(t) = \mathscr{D}(t). \end{aligned} \quad (30.32)$$

Remarquons que puisque $\mathscr{D}(t)$ et $\mathscr{E}(t)$ sont des matrices symétriques, on a d'après (24) et (32) les relations suivantes:

$$\mathscr{V}(t)=[\mathscr{E}(t)\widetilde{C}^*(t)+B(t)Q(t)B^*(t)C^*]\mathscr{D}^{-1}(t), \qquad (30.33)$$

$$\mathscr{V}^*(t)=\mathscr{D}^{-1}(t)[\widetilde{C}(t)\mathscr{E}(t)+CB(t)Q(t)B^*(t)], \qquad (30.34)$$

$$\begin{aligned}\mathscr{V}^*(t)\mathscr{D}(t)\mathscr{V}(t)=\mathscr{E}(t)\widetilde{C}^*(t)\mathscr{D}^{-1}(t)\widetilde{C}(t)\mathscr{E}(t)+\\+\mathscr{E}(t)\widetilde{C}^*(t)\mathscr{D}^{-1}(t)CB(t)Q(t)B^*(t)+\\+B(t)Q(t)B^*(t)C^*\mathscr{D}^{-1}(t)\widetilde{C}(t)\mathscr{E}(t)+\\+B(t)Q(t)B^*(t)C^*\mathscr{D}^{-1}(t)CB(t)Q(t)B^*(t).\end{aligned} \qquad (30.35)$$

En portant l'expression (35) dans l'équation (25), on réduit cette dernière à la forme

$$\frac{d\mathscr{E}}{dt}=\widetilde{A}(t)\mathscr{E}+\mathscr{E}\widetilde{A}^*(t)-\mathscr{E}\widetilde{C}^*(t)\widetilde{R}^{-1}(t)\widetilde{C}(t)\mathscr{E}+\\+\widetilde{B}(t)\widetilde{Q}(t)\widetilde{B}^*(t), \quad \mathscr{E}(t_0)=\mathscr{T}. \qquad (30.36)$$

L'équation différentielle (30) du filtre optimal s'écrit en vertu de (33) et (32)

$$\frac{d\gamma}{dt}=[\widetilde{A}(t)-\mathscr{E}(t)\widetilde{C}^*(t)\widetilde{R}^{-1}(t)\widetilde{C}(t)]\gamma+\mathscr{N}(t)z(t), \qquad (30.37)$$

$$\gamma(t_0)=[\Gamma C^*\mathscr{U}^{-1}-\mathscr{V}(t_0)]z(t_0).$$

L'équation différentielle de Riccati (36) ne se distingue de (28.80) que par le remplacement des paramètres $A(t)$, $B(t)$, $C(t)$, $Q(t)$ et $R(t)$ par $\widetilde{A}(t)$, $\widetilde{B}(t)$, $\widetilde{C}(t)$, $\widetilde{Q}(t)$ et $\widetilde{R}(t)$. Cette même correspondance caractérise les équations différentielles homogènes qui se forment à partir des équations (37) et (28.82) avec $z(t)\equiv 0$. C'est pourquoi les résultats du § 29 s'étendent également à notre problème. Bien entendu, dans ces conditions, les fonctions $A(t)$, $B(t)$, $C(t)$, $Q(t)$ et $R(t)$ doivent être remplacées respectivement par $\widetilde{A}(t)$, $\widetilde{B}(t)$, $\widetilde{C}(t)$, $\widetilde{Q}(t)$ et $\widetilde{R}(t)$.

§ 31. Filtrage optimal à valeurs discontinues du temps

Considérons le système régi par les équations aux différences finies

$$x(t_{n+1})=\Phi(t_{n+1},t_n)x(t_n)+\Gamma(t_n)w_n, \qquad n=0,1,2,\ldots$$

$$x(t_0)=s, \qquad (31.1)$$

$$z_n=H_nx(t_n)+v_n.$$

Ici x est le vecteur d'état de dimension n; $\Phi(t_{n+1}, t_n)$, la matrice $n \times n$; $\Gamma(t_n)$, la matrice $n \times r$; z_n, le vecteur de dimension m; H_n, la matrice $m \times n$; w_n et v_n, les suites vectorielles de dimension r et de dimension m respectivement des variables aléatoires gaussiennes réciproquement indépendantes à moyennes nulles; s, la variable aléatoire vectorielle gaussienne de dimension n à moyenne nulle; de plus s, w_n et v_n sont réciproquement indépendantes.

Les matrices de corrélation des suites discontinues w_n et v_n sont

$$\begin{aligned} \mathrm{M}[w_n w_k^*] &= Q_n \delta_{nk}, \\ \mathrm{M}[v_n v_k^*] &= R_n \delta_{nk}, \\ \mathrm{M}[w_n v_k^*] &= 0. \end{aligned} \tag{31.2}$$

Ici Q_n est la matrice symétrique définie non négative $r \times r$; R_n, la matrice symétrique définie positive $m \times m$; δ_{nk}, le symbole de Kronecker

$$\delta_{nk} = \begin{cases} 1 & \text{pour } n = k \\ 0 & \text{pour } n \neq k. \end{cases}$$

Désignons par $\hat{x}(t_n \mid j)$ l'espérance mathématique conditionnelle de $x(t_n)$ pour $z_0, z_1, \ldots, z_j$ donnés:

$$\begin{aligned} \hat{x}(t_n \mid j) &= \mathrm{M}[x(t_n) \mid z_0, z_1, \ldots, z_j], \\ &\cdots\cdots\cdots \\ \hat{x}(t_n \mid n) &= \mathrm{M}[x(t_n) \mid z_0, z_1, \ldots, z_j, \ldots, z_n]. \end{aligned} \tag{31.3}$$

De même que dans le cas continu (§§ 27, 28), d'après les données d'observation de $z_0, z_1, \ldots, z_n$, $\hat{x}(t_n \mid n)$ est la meilleure estimation quadratique moyenne de $x(t_n)$.

Théorème. *Supposons que $R_n \geqslant \alpha E$, où $\alpha > 0$, et E la matrice unité $m \times m$. Alors, les relations récurrentes suivantes ont lieu*

$$\hat{x}(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n)\, \hat{x}(t_n \mid n), \tag{31.4}$$

$$\begin{gathered} \hat{x}(t_n \mid n) = \hat{x}(t_n \mid n-1) + A_n [z_n - H_n \hat{x}(t_n \mid n-1), \\ \hat{x}(t_0 \mid -1) = \hat{x}(t_0) = 0, \end{gathered} \tag{31.5}$$

$$A_n = P(t_n \mid n-1) H_n^* [H_n P(t_n \mid n-1) H_n^* + R_n]^{-1}, \tag{31.6}$$

$$P(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n) P(t_n \mid n) \Phi^*(t_{n+1}, t_n) + \Gamma(t_n) Q_n \Gamma^*(t_n), \tag{31.7}$$

$$\begin{gathered} P(t_n \mid n) = [E - A_n H_n] P(t_n \mid n-1) [E - A_n H_n]^* + \\ + A_n R_n A_n^* = P(t_n \mid n-1) - A_n H_n P(t_n \mid n-1), \\ P(t_0 \mid -1) = P(t_0) = \mathrm{M}[ss^*]. \end{gathered} \tag{31.8}$$

Ici

$$P(t_n \mid n) = \mathrm{M}[\tilde{x}(t_n \mid n)\tilde{x}^*(t_n \mid n)],$$

$$P(t_n \mid n-1) = \mathrm{M}[\tilde{x}(t_n \mid n-1)\tilde{x}^*(t_n \mid n-1)], \qquad (31.9)$$

$$\tilde{x}(t_n \mid j) = x(t_n) - \hat{x}(t_n \mid j). \qquad (31.10)$$

Les relations récurrentes (7) *et* (8) *entraînent que*

$$P(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n) P(t_n \mid n-1) \Phi^*(t_{n+1}, t_n) - \\ -\Phi(t_{n+1}, t_n) P(t_n \mid n-1) H_n^* [H_n P(t_n \mid n-1) H_n^* + R_n]^{-1} \times \\ \times H_n P(t_n \mid n-1) \Phi^*(t_{n+1}, t_n) + \\ + \Gamma(t_n) Q_n \Gamma^*(t_n), \qquad (31.11)$$

$$P(t_0 \mid -1) = P(t_0) = \mathrm{M}[ss^*].$$

La relation (11) *est une équation matricielle aux différences de Riccati. Cette équation se transforme en équation différentielle matricielle de Riccati à la limite pour* $t_n - t_{n-1} \to 0$.

D'une façon analogue, les relations récurrentes (4) *et* (5) *et la formule* (6) *impliquent que*

$$\hat{x}(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n)\hat{x}(t_n \mid n-1) + \\ + \Phi(t_{n+1}, t_n) P(t_n \mid n-1) H_n^* [H_n P(t_n \mid n-1) H_n^* + R_n]^{-1} \times \\ \times [z_n - H_n \hat{x}(t_n \mid n-1)],$$

$$\hat{x}(t_0 \mid -1) = \hat{x}(t_0) = 0. \qquad (31.12)$$

La relation (12) *représente une équation vectorielle aux différences qui décrit le filtre optimal.*

D é m o n s t r a t i o n. Considérons d'abord certaines propriétés des répartitions de probabilités conditionnelles des variables aléatoires vectorielles gaussiennes.

1°. Soit k le vecteur de dimension $n + m$

$$k = \begin{bmatrix} a \\ b \end{bmatrix}, \qquad (31.13)$$

où le vecteur a de dimension n et le vecteur b de dimension m sont des variables aléatoires vectorielles gaussiennes à moyennes nulles. Supposons encore que $\mathrm{M}[bb^*] > 0$.

Introduisons les notations

$$N = \mathrm{M}[kk^*], \qquad N_a = \mathrm{M}[aa^*], \\ N_{ab} = \mathrm{M}[ab^*], \qquad N_{ba} = \mathrm{M}[ba^*], \qquad (31.14) \\ N_b = \mathrm{M}[bb^*].$$

On a alors

$$N = \begin{bmatrix} N_a & N_{ab} \\ N_{ba} & N_b \end{bmatrix}. \qquad (31.15)$$

Désignons par c le vecteur

$$c = a - N_{ab}N_b^{-1}b. \tag{31.16}$$

Constatons que

$$N_c = \mathrm{M}\,[cc^*] = \mathrm{M}\,[(a - N_{ab}N_b^{-1}b)\ (a - N_{ab}N_b^{-1}b)^*] =$$
$$= \mathrm{M}\,[aa^*] - \mathrm{M}\,[ab^*]\,N_b^{-1}N_{ba} - N_{ab}N_b^{-1}\mathrm{M}\,[ba^*] +$$
$$+ N_{ab}N_b^{-1}\mathrm{M}\,[bb^*]\,N_b^{-1}N_{ba} =$$
$$= N_a - N_{ab}N_b^{-1}N_{ba} - N_{ab}N_b^{-1}N_{ba} + N_{ab}N_b^{-1}N_{ba},$$

ou

$$N_c = N_a - N_{ab}N_b^{-1}N_{ba}. \tag{31.17}$$

Cherchons maintenant

$$\det N = \det \begin{bmatrix} N_a & N_{ab} \\ N_{ba} & N_b \end{bmatrix}.$$

En multipliant la deuxième ligne de la matrice N à gauche par la matrice $-N_{ab}N_b^{-1}$ et en ajoutant les éléments de la ligne obtenue aux éléments de la première ligne de la matrice N, on trouve

$$\det N = \det \begin{bmatrix} N_a & N_{ab} \\ N_{ba} & N_b \end{bmatrix} = \det \begin{bmatrix} N_a - N_{ab}N_b^{-1}N_{ba} & 0 \\ N_{ba} & N_b \end{bmatrix} =$$
$$= \det (N_a - N_{ab}N_b^{-1}N_{ba}) \det N_b,$$

ou, d'après (17),

$$\det N = \det N_c \det N_b. \tag{31.18}$$

Conformément à (16), on a la transformation linéaire

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E & N_{ab}N_b^{-1} \\ 0 & E \end{bmatrix} \begin{bmatrix} c \\ b \end{bmatrix}. \tag{31.19}$$

Comme le jacobien de la transformation (19) vaut l'unité, la densité de probabilité commune des vecteurs aléatoires a et b s'écrit

$$f(a,\ b) = (2\pi)^{-\frac{n+m}{2}} (\det N)^{-\frac{1}{2}} e^{-\frac{1}{2}(c^*N_c^{-1}\,c + b^*N_b^{-1}b)}. \tag{31.20}$$

La densité de probabilité du vecteur aléatoire b est

$$f(b) = (2\pi)^{-\frac{m}{2}} (\det N_b)^{-\frac{1}{2}} e^{-\frac{1}{2}b^*N_b^{-1}b}. \tag{31.21}$$

La densité de probabilité conditionnelle du vecteur aléatoire a pour b donné:

$$f(a \mid b) = \frac{f(a,\ b)}{f(b)} = (2\pi)^{-\frac{n}{2}} (\det N_c)^{-\frac{1}{2}} e^{-\frac{1}{2}c^*N_c^{-1}c},$$

ou, en vertu de (16),

$$f(a \mid b) = (2\pi)^{-\frac{n}{2}} (\det N_c)^{-\frac{1}{2}} e^{-\frac{1}{2}(a - N_{ab}N_b^{-1}b)^*N_c^{-1}(a - N_{ab}N_b^{-1}b)}. \tag{31.22}$$

Il suit de la relation (22) que l'espérance mathématique conditionnelle du vecteur a pour b donné s'écrit

$$\mathrm{M}\,[a \mid b] = N_{ab}N_b^{-1}b \tag{31.23}$$

ou, d'après (14),

$$\mathrm{M}\,[a \mid b] = \mathrm{M}\,[ab^*]\,\{\mathrm{M}\,[bb^*]\}^{-1}\,b. \tag{31.24}$$

Comme d'après (16)

$$a - \mathrm{M}\,[a \mid b] = a - N_{ab}N_b^{-1}b = c,$$

il vient

$$\mathrm{M}\,[(a - \mathrm{M}\,[a \mid b])\,(a - \mathrm{M}\,[a \mid b])^* \mid b] = N_c = N_a - N_{ab}N_b^{-1}N_{ba}. \tag{31.25}$$

Notons que l'indépendance de la matrice de corrélation conditionnelle (25) par rapport à b est une propriété importante des processus aléatoires gaussiens. Pour d'autres processus aléatoires, cette relation a lieu, ce qui complique nettement le problème général d'estimation optimale.

2°. Montrons maintenant l'indépendance des vecteurs aléatoires c et b. D'après (16),

$$\mathrm{M}\,[cb^*] = \mathrm{M}\,[(a - N_{ab}N_b^{-1}b)\,b^*] =$$
$$= \mathrm{M}\,[ab^*] - N_{ab}N_b^{-1}\mathrm{M}\,[bb^*] = N_{ab} - N_{ab}N_b^{-1}N_b = 0.$$

Ainsi,

$$\mathrm{M}\,[(a - \mathrm{M}\,[a \mid b])\,b^*] = \mathrm{M}\,[cb^*] = 0, \tag{31.26}$$

et, donc, les vecteurs aléatoires $c = a - \mathrm{M}\,[a \mid b]$ et b sont indépendants.

3°. Soient le vecteur a de dimension n, le vecteur b de dimension m et le vecteur e de dimension r, variables aléatoires vectorielles gaussiennes à moyennes nulles ; de plus,

$$\mathrm{M}\,[bb^*] > 0, \qquad \mathrm{M}\,[ee^*] > 0, \qquad \mathrm{M}\,[be^*] = 0. \tag{31.27}$$

Désignons par l le vecteur de dimension $(m + r)$

$$l = \begin{bmatrix} b \\ e \end{bmatrix}. \tag{31.28}$$

Alors, d'après (24),

$$\mathrm{M}\,[a \mid l] = \mathrm{M}\,[al^*]\,\{\mathrm{M}\,[ll^*]\}^{-1}\,l. \tag{31.29}$$

Comme $M\,[be^*] = 0$, il vient

$$\mathrm{M}\,[ll^*] = \begin{bmatrix} \mathrm{M}\,[bb^*] & 0 \\ 0 & \mathrm{M}\,[ee^*] \end{bmatrix}, \quad \{\mathrm{M}\,[ll^*]\}^{-1} = \begin{bmatrix} \{\mathrm{M}\,[bb^*]\}^{-1} & 0 \\ 0 & \{\mathrm{M}\,[ee^*]\}^{-1} \end{bmatrix},$$
$$\mathrm{M}\,[al^*] = [\mathrm{M}\,[ab^*], \ \mathrm{M}\,[ae^*]]. \tag{31.30}$$

Il en résulte que

$$\mathrm{M}\,[al^*]\,\{\mathrm{M}\,[ll^*]\}^{-1}\,l = \mathrm{M}\,[ab^*]\,\{\mathrm{M}\,[bb^*]\}^{-1}\,b + \mathrm{M}\,[ae^*]\,\{\mathrm{M}\,[ee^*]\}^{-1}\,e \tag{31.31}$$

ou, en vertu de (24) et (29),

$$\mathrm{M}\,[a \mid b, \ e] = \mathrm{M}\,[a \mid b] + \mathrm{M}\,[a \mid e]. \tag{31.32}$$

Proposons-nous maintenant de démontrer le théorème. La relation (4) se déduit de l'équation initiale (1) du fait que

$$M[w_n] = 0.$$

Désignons par $\tilde{z}_n$ le vecteur

$$\tilde{z}_n = z_n - H_n \hat{x}(t_n \mid n-1). \quad (31.33)$$

Puisque d'après (3), $\hat{x}(t_n \mid n-1)$ n'est défini que d'après $z_0, z_1, \ldots, z_{n-1}$ donnés et présente une combinaison linéaire de ces vecteurs, on a la relation

$$M[x(t_n) \mid z_0, z_1, \ldots, z_n] = M[x(t_n) \mid z_0, \ldots, z_{n-1}, \tilde{z}_n]. \quad (31.34)$$

D'après (1) et (3), le vecteur $\tilde{z}_n$ se met sous la forme

$$\tilde{z}_n = H_n [x(t_n) - \hat{x}(t_n \mid n-1)] + v_n =$$
$$= H_n \{x(t_n) - M[x(t_n) \mid z_0, z_1, \ldots, z_{n-1}]\} + v_n.$$

Comme il résulte des relations (26) et (2), les vecteurs aléatoires $\tilde{z}_n$ et $(z_0, z_1, \ldots, z_{n-1})$ sont indépendants. Par suite, conformément à (32):

$$M[x(t_n) \mid z_0, \ldots, z_{n-1}, \tilde{z}_n] =$$
$$= M[x(t_n) \mid z_0, \ldots, z_{n-1}] + M[x(t_n) \mid \tilde{z}_n]. \quad (31.35)$$

En vertu des relations (3), (34) et (24), la formule (35) peut être ramenée à la forme

$$\hat{x}(t_n \mid n) = \hat{x}(t_n \mid n-1) + M[x(t_n) \mid \tilde{z}_n] =$$
$$= \hat{x}(t_n \mid n-1) + M[x(t_n) \tilde{z}_n^*] \{M[\tilde{z}_n \tilde{z}_n^*]\}^{-1} \tilde{z}_n. \quad (31.36)$$

Si l'on tient compte que d'après (9)

$$P(t_n \mid n-1) = M[\tilde{x}(t_n \mid n-1) \tilde{x}^*(t_n \mid n-1)].$$

on obtient

$$M[\tilde{z}_n \tilde{z}_n^*] = H_n P(t_n \mid n-1) H_n^* + R_n, \quad (31.37)$$

$$M[x(t_n) \tilde{z}_n^*] = M[x(t_n) \tilde{x}^*(t_n \mid n-1)] H_n^* = P(t_n \mid n-1) H_n^*. \quad (31.38)$$

La formule (37) s'obtient de la façon suivante. D'après (33), (1) et (10),

$$\tilde{z}_n = H_n x(t_n) + v_n - H_n \hat{x}(t_n \mid n-1)$$

ou

$$\tilde{z}_n = H_n \tilde{x}(t_n \mid n-1) + v_n. \quad (31.39)$$

On voit de (12) que $\hat{x}(t_n \mid n-1)$, et, par conséquent, $\tilde{x}(t_n \mid n-1)$, sont indépendants de z_n, et donc de v_n. Donc,

$$\mathbf{M}[\tilde{z}_n\tilde{z}_n^*] = \mathbf{M}\{H_n\tilde{x}(t_n \mid n-1)\,\tilde{x}^*(t_n \mid n-1)\,H_n^*\} + \mathbf{M}[v_nv_n^*], \quad (31.40)$$

d'où l'on tire précisément la formule (37).

Pour obtenir la formule (38), notons que puisque

$$\mathbf{M}[x(t_n)\,v_n^*] = 0,$$

il vient, d'après (39),

$$\mathbf{M}[x(t_n)\,\tilde{z}_n^*] = \mathbf{M}[x(t_n)\,\tilde{x}^*(t_n \mid n-1)\,H_n^*]. \quad (31.41)$$

En vertu de (10), l'expression (41) devient

$$\mathbf{M}[x(t_n)\,\tilde{z}_n^*] = \mathbf{M}\{[\tilde{x}(t_n \mid n-1) + \hat{x}(t_n \mid n-1)]\,\tilde{x}^*(t_n \mid n-1)\,H_n^*\},$$

ou

$$\mathbf{M}[x(t_n)\,\tilde{z}_n^*] = P(t_n \mid n-1)\,H_n^* + \mathbf{M}[\hat{x}(t_n \mid n-1)\,\tilde{x}^*(t_n \mid n-1)]\,H_n^*. \quad (31.42)$$

En raison de la relation (27.35) qui traduit le contenu du lemme sur la projection orthogonale,

$$\mathbf{M}[\hat{x}(t_n \mid n-1)\,\tilde{x}^*(t_n \mid n-1)] = 0. \quad (31.43)$$

Les relations (42) et (43) conduisent à la formule (38).

Les formules (36), (37), (38) et (33) justifient la récurrence (5))

Pour obtenir la relation récurrente (8), remarquons que d'après (10), (5) et (33)

$$\tilde{x}(t_n \mid n) = x(t_n) - \hat{x}(t_n \mid n) = x(t_n) - \hat{x}(t_n \mid n-1) - A_n\tilde{z}_n,$$

ou

$$\tilde{x}(t_n \mid n) = \tilde{x}(t_n \mid n-1) - A_n\tilde{z}_n. \quad (31.44)$$

Il en résulte que

$$\mathbf{M}[\tilde{x}(t_n \mid n)\,\tilde{x}^*(t_n \mid n)] = \\ = \mathbf{M}\{[\tilde{x}(t_n \mid n-1) - A_n\tilde{z}_n]\,[\tilde{x}(t_n \mid n-1) - A_n\tilde{z}_n]^*\}$$

ou, d'après (9),

$$P(t_n \mid n) = P(t_n \mid n-1) - A_n\mathbf{M}[\tilde{z}_n\tilde{x}^*(t_n \mid n-1)] - \\ - \mathbf{M}[\tilde{x}(t_n \mid n-1)\,\tilde{z}_n^*]\,A_n^* + A_n\mathbf{M}[\tilde{z}_n\tilde{z}_n^*]\,A_n^*. \quad (31.45)$$

Comme $\tilde{x}(t_n \mid n-1)$ et v_n sont indépendants, (39) et (9) entraînent

$$A_n\mathbf{M}[\tilde{z}_n\tilde{x}^*(t_n \mid n-1)] = A_n\mathbf{M}[H_n\tilde{x}(t_n \mid n-1)\,\tilde{x}^*(t_n \mid n-1)] = \\ = A_nH_nP(t_n \mid n-1), \quad (31.46)$$

$$\mathbf{M}[\tilde{x}(t_n \mid n-1)\,\tilde{z}_n^*]\,A_n^* = \mathbf{M}[\tilde{x}(t_n \mid n-1)\,\tilde{x}^*(t_n \mid n-1)]\,H_n^*A_n^* = \\ = P(t_n \mid n-1)\,H_n^*A_n^*. \quad (31.47)$$

D'après (37),

$$A_n \mathrm{M}\,[\tilde{z}_n \tilde{z}_n^*]\, A_n^* = A_n\,[H_n P\,(t_n \mid n-1)\, H_n^* + R_n]\, A_n^*,$$

d'où, conformément à (6), on a

$$A_n \mathrm{M}\,[\tilde{z}_n \tilde{z}_n^*]\, A_n^* = P\,(t_n \mid n-1)\, H_n^* A_n^*. \qquad (31.48)$$

En portant les expressions (46), (47) et (48) dans la relation (45), on obtient la relation (8) :

$$P\,(t_n \mid n) = P\,(t_n \mid n-1) - A_n H_n P\,(t_n \mid n-1).$$

Ainsi nous avons justifié la relation récurrente (8).

§ 32. Application de la programmation dynamique à la théorie de la commande optimale des systèmes stochastiques

1. Equation différentielle stochastique d'Ito. Considérons le système décrit par l'équation différentielle stochastique vectorielle

$$dx = f(t, x)dt + G\,(t, x)dw\,(t),\ t_0 \leqslant t \leqslant T. \qquad (32.1)$$

Ici x est le vecteur de dimension n ; $f\,(t, x)$, le vecteur fonction de dimension n ; $w\,(t)$, le processus séparable de Wiener de dimension m ; $G\,(t, x)$, la matrice $n \times m$.

Puisque

$$\zeta = \frac{dw}{dt} \qquad (32.2)$$

est un bruit blanc gaussien de dimension m à moyenne nulle, l'équation (1) se met quelquefois sous la forme

$$\frac{dx}{dt} = f\,(t, x) + G\,(t, x)\,\zeta\,(t). \qquad (32.3)$$

La matrice de corrélation du processus aléatoire $\zeta\,(t)$ a pour expression

$$\mathrm{cov}\,[\zeta\,(t);\ \zeta\,(\tau)] = \mathrm{M}\,[\zeta\,(t)\,\zeta^*\,(\tau)] = Q\,(t)\,\delta\,(t-\tau). \qquad (32.4)$$

Ici $Q\,(t)$ est la matrice symétrique définie non négative $m \times m$; cette matrice peut donc se mettre sous la forme

$$Q\,(t) = S\,(t)\,S^*\,(t), \qquad (32.5)$$

où $S\,(t)$ est la matrice singulière réelle $m \times m$.

Introduisons les notations:

$$dw\,(t) = S\,(t)\,d\tilde{w}\,(t), \qquad (32.6)$$

$$\zeta\,(t) = S\,(t)\,\tilde{\zeta}\,(t). \qquad (32.7)$$

L'équation (1) se met alors sous la forme

$$dx = f\,(t, x)\,dt + g\,(t, x)\,d\tilde{w}\,(t), \qquad t_0 \leqslant t \leqslant T, \qquad (32.8)$$

où

$$g\,(t, x) = G\,(t, x)\,S\,(t). \qquad (32.9)$$

D'une façon analogue, l'équation (3) devient

$$\frac{dx}{dt} = f(t, x) + g(t, x)\tilde{\zeta}(t). \qquad (32.10)$$

Ici $\tilde{\zeta}(t)$ est le bruit blanc gaussien de dimension m à moyenne nulle et à matrice de corrélation

$$\operatorname{cov}[\tilde{\zeta}(t); \tilde{\zeta}(\tau)] = \mathrm{M}[\tilde{\zeta}(t)\tilde{\zeta}^*(\tau)] = E\delta(t-\tau), \qquad (32.11)$$

où E est la matrice unité $m \times m$.

L'équation différentielle stochastique (8) (et, par suite, l'équation (10)), est équivalente à l'équation intégrale stochastique

$$x(t) = x(t_0) + \int_{t_0}^{t} f(\tau, x(\tau))\,d\tau + \int_{t_0}^{t} g(\tau, x(\tau))\,d\tilde{w}(\tau). \qquad (32.12)$$

On suppose que les fonctions $f(t, x)$ et $g(t, x)$ sont mesurables par rapport à t et à x pour $t \in [t_0, T]$ et vérifient les conditions

$$\|f(t, x)\| + \|g(t, x)\| \leqslant k(1 + \|x\|), \qquad t \in [t_0, T],$$
$$\|f(t, x) - f(t, y)\| + \|g(t, x) - g(t, y)\| \leqslant$$
$$\leqslant k\|x - y\|, \qquad t \in [t_0, T],$$

où k est la constante correspondante. Dans l'équation (12), $x(t_0)$ est une variable aléatoire quelconque indépendante de $d\tilde{w}$. Sous ces conditions, l'équation intégrale stochastique (12) définit le processus stochastique unique $X = \{x(t);\ t \in [t_0, T]\}$ qui est un processus *markovien*.

2. Processus de diffusion. Formule d'intégration d'Ito. L'équation différentielle stochastique (8) définit la famille des processus markoviens qui dépend du paramètre aléatoire $x(t_0)$. En général, nous désignerons par X un processus fixé mais pris arbitrairement dans cette famille. Le processus markovien X est un processus *de diffusion*.

Soit la fonction scalaire $\psi(t, x)$, où x est le vecteur de dimension n qui vérifie l'équation différentielle stochastique (8). Supposons que $\psi(t, x)$, $\frac{\partial\psi}{\partial t}$, $\frac{\partial\psi}{\partial x_k}$, $k = 1, \ldots, n$, $\frac{\partial^2\psi}{\partial x_j\,\partial x_k}$, $j, k = 1, \ldots, n$, sont des fonctions continues par rapport à leurs variables.

Désignons par ψ_x le vecteur de dimension n

$$\psi_x = \begin{bmatrix} \partial\psi/\partial x_1 \\ \cdots \\ \partial\psi/\partial x_n \end{bmatrix}, \qquad (32.13)$$

et par ψ_{xx}, la matrice symétrique $n \times n$

$$\psi_{xx} = \begin{bmatrix} \frac{\partial^2\psi}{\partial x_1^2} & \frac{\partial^2\psi}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2\psi}{\partial x_1\,\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2\psi}{\partial x_1\,\partial x_n} & \frac{\partial^2\psi}{\partial x_2\,\partial x_n} & \cdots & \frac{\partial^2\psi}{\partial x_n^2} \end{bmatrix}. \qquad (32.14)$$

Désignons par $\sigma(t, x)$ la matrice $n \times n$

$$\sigma(t, x) = g(t, x)\, g^*(t, x). \qquad (32.15)$$

La relation auxiliaire

$$\frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_{jk}(t, x) \frac{\partial^2\psi}{\partial x_j\,\partial x_k} = \frac{1}{2} \operatorname{Sp}[g^*(t, x)\,\psi_{xx} g(t, x)] \qquad (32.16)$$

se déduit du fait que d'après (41.50) et (15)

$$\operatorname{Sp}[g^*(t, x)\,\psi_{xx} g(t, x)] = \operatorname{Sp}[\psi_{xx} g(t, x)\, g^*(t, x)] = \\ = \operatorname{Sp}[\psi_{xx}\sigma(t, x)] = \operatorname{Sp}[\sigma(t, x)\,\psi_{xx}]. \qquad (32.17)$$

Comme d'après (41.50), (9) et (5),

$$\operatorname{Sp}[g^*(t, x)\,\psi_{xx} g(t, x)] = \operatorname{Sp}[\psi_{xx} g(t, x)\, g^*(t, x)] = \\ = \operatorname{Sp}[\psi_{xx} G(t, x)\, S(t)\, S^*(t)\, G^*(t, x)] = \\ = \operatorname{Sp}[\psi_{xx} G(t, x)\, Q(t)\, G^*(t, x)], \qquad (32.18)$$

la relation (16) devient

$$\frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_{jk}(t, x) \frac{\partial^2\psi}{\partial x_j\,\partial x_k} = \frac{1}{2} \operatorname{Sp}[\psi_{xx} G(t, x)\, Q(t)\, G^*(t, x)]. \qquad (32.19)$$

On sait [49] que l'*opérateur infinitésimal* $\mathfrak{D}$ du processus markovien aléatoire X, défini par l'équation intégrale stochastique (12), est de la forme $\mathfrak{D} = \mathcal{L} + \frac{\partial}{\partial t}$, où l'opérateur $\mathcal{L}$ peut être déterminé à l'aide de la relation

$$\mathcal{L}\psi(t, x) = \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_{jk}(t, x) \frac{\partial^2\psi}{\partial x_j\,\partial x_k} + f^*(t, x)\,\psi_x, \qquad (32.20)$$

où la matrice $\sigma(t)$ du type $n \times n$ et le vecteur ψ_x de dimension n sont déterminés par les expressions (15) et (13).

Conformément à (16), on a pour l'expression (20),

$$\mathcal{L}\psi(t, x) = \frac{1}{2} \operatorname{Sp}[g^*(t, x)\,\psi_{xx} g(t, x)] + f^*(t, x)\,\psi_x \qquad (32.21)$$

ou, d'après (19),

$$\mathcal{L}\psi(t, x) = \frac{1}{2} \operatorname{Sp}[\psi_{xx} G(t, x)\, Q(t)\, G^*(t, x)] + f^*(t, x)\,\psi_x. \qquad (32.22)$$

De plus, comme l'a montré Ito [25],

$$\psi(t, x(t)) = \psi(s, x(s)) + \int_s^t d\psi(\tau, x(\tau)) \text{ pour } t \geqslant s, \tag{32.23}$$

où $d\psi$ est la *différentielle stochastique d'Ito* de la forme

$$d\psi(\tau, x(\tau)) = [\mathscr{L}\psi(\tau, x(\tau)) + \psi_\tau(\tau, x(\tau))]\, d\tau + \\ + \psi_x^*(\tau, x(\tau))\, g(\tau, x(\tau))\, d\tilde{w}(\tau). \tag{32.24}$$

Ici

$$\psi_\tau = \frac{\partial\psi(\tau, x)}{\partial\tau}. \tag{32.25}$$

On en déduit la *formule d'intégration d'Ito* [25],

$$\mathbf{M}_{s,x}[\psi(t, x(t))] = \\ = \psi(s, x) + \mathbf{M}_{s,x}\left\{\int_s^t [\mathscr{L}\psi(\tau, x(\tau)) + \psi_\tau(\tau, x(\tau))]\, d\tau\right\}. \tag{32.26}$$

Le symbole $\mathbf{M}_{s,x}$ désigne ici l'espérance mathématique sous la condition qu'à l'instant s le processus markovien $x(t)$ «sort» du point x.

3. Commande optimale des systèmes stochastiques. Equation de Bellman. Considérons le système commandé décrit par l'équation différentielle stochastique

$$dx = [A(t)x + L(t)u]\,dt + g(t, x)\,d\tilde{w}(t), \quad x(t_0) = x_0. \tag{32.27}$$

Ici x est le vecteur de dimension n ; $A(t)$, la matrice $n \times n$; u, le vecteur de commande de dimension p ; $L(t)$, la matrice $n \times p$; $\tilde{w}(t)$, le processus séparable de Wiener de dimension m défini dans ce qui précède par la relation (6) ; $g(t, x)$, la matrice $n \times m$ définie par la relation (9).

On demande de choisir la commande u telle qu'elle minimise la fonctionnelle

$$J(u) = \mathbf{M}\left[b(T, x(T)) + \int_{t_0}^T \Lambda(t, x(t), u(t))\,dt\right], \tag{32.28}$$

où $\Lambda(t, x, u)$ est une fonction scalaire positive de ses variables. La grandeur T est fixée ; la valeur de $x(T)$ n'est pas fixée à l'avance. Nous avons donc un problème à extrémité libre de trajectoire et à temps fixé.

Désignons par $\mathcal{U}$ la classe des fonctions $\varphi(t, x)$ qui jouissent des propriétés suivantes: $\varphi(t, x)$ est continue par morceaux par rapport à t pour tout x fixé,

$$\|\varphi(t, x)\| \leqslant k(1 + \|x\|), \quad (t, x) \in [t_0, T] \times R^n \tag{32.29}$$

et $\varphi(t, x)$ satisfait uniformément à la condition de Lipschitz par rapport à x dans $[t_0, T] \times R^n$.

Considérons comme *admissibles* les commandes

$$u(t) = \varphi(t, x(t)) \tag{32.30}$$

et notons-les $u \in \mathcal{U}$.

Supposons également que x_0 est un vecteur constant (non aléatoire).

Sous les conditions qui viennent d'être énoncées, l'équation différentielle stochastique (27) définit dans l'intervalle (t_0, T) le processus de diffusion X_φ. L'opérateur différentiel $\mathcal{L}_u$ de ce processus se définit, d'une façon analogue à (21), par la relation

$$\mathcal{L}_u V(t, x) = \frac{1}{2}\operatorname{Sp}[g^*(t, x) V_{xx} g(t, x)] + \\ + [A(t)x + L(t)u]^* V_x, \tag{32.31}$$

où u doit être remplacée par $\varphi(t, x)$.

L e m m e. *Supposons qu'il existe une commande $\varphi^0(t, x)$ et une fonction scalaire $V(t, x)$, $t \in [t_0, T]$ qui jouissent des propriétés:*

1°. *V, V_t, V_x, V_{xx} sont continues.*

2°. *Pour tout x et tout $t \in [t_0, T]$,*

$$0 = V_t(t, x) + \mathcal{L}_{\varphi^0} V(t, x) + \Lambda(t, x, \varphi^0(t, x)), \tag{32.32}$$

$$0 \leqslant V_t(t, x) + \mathcal{L}_u V(t, x) + \Lambda(t, x, u). \tag{32.33}$$

3°. *Pour $t = T$,*

$$V(T, x(T)) = b(T, x(T)). \tag{32.34}$$

Alors la commande $u^0 = \varphi^0(t, x)$ est optimale.

D é m o n s t r a t i o n. Soient φ une commande admissible arbitraire et $x(t)$ et $x^0(t)$, les trajectoires des processus X_φ et X_{φ^0} respectivement.

La formule d'intégration d'Ito (26) devient alors

$$\mathrm{M}_{t,x}[V(T, x(T))] = \\ = V(t, x) + \mathrm{M}_{t,x}\left\{\int_t^T [\mathcal{L}_\varphi V(\tau, x(\tau)) + V_\tau(\tau, x(\tau))]\, d\tau\right\}. \tag{33.35}$$

Puisque d'après (34), $V(T, x(T)) = b(T, x(T))$, les conditions (32) étant remplies, la relation (35) s'écrit

$$V(t,x) = -\mathbf{M}_{t,x}\left\{\int_t^T [\mathcal{L}_{\varphi^0} V(\tau, x^0(\tau)) + V'_\tau(\tau, x^0(\tau))]d\tau - b(T, x^0(T))\right\} =$$

$$= \mathbf{M}_{t,x}\left[b(T, x^0(T)) + \int_t^T \Lambda(\tau, x^0(\tau)), \varphi^0(\tau, x^0(\tau))\,d\tau\right]. \quad (32.36)$$

D'une façon analogue, lorsque la condition (33) est remplie, la relation (35) devient

$$V(t, x) = -\mathbf{M}_{t,x}\left\{\int_t^T [\mathcal{L}_\varphi V(\tau, x(\tau)) + V_\tau(\tau, x(\tau))]\,d\tau - \right.$$

$$\left. - b(T, x(T))\right\} \leqslant \mathbf{M}_{t,x}\left[b(T, x(T)) + \int_t^T \Lambda(\tau, x(\tau), \varphi(\tau, x(\tau))\,d\tau\right]. \quad (32.37)$$

En posant dans les relations (36) et (37) $t = t_0$, $x = x_0$ on obtient, conformément à (28), la relation

$$V(t_0, x_0) = J(u^0) \leqslant J(u). \quad (32.38)$$

Il résulte de (38) que la commande $u^0 = \varphi^0(t, x)$ vérifiant les conditions (32), (33) et (34) est optimale. Le lemme est démontré.

Les conditions (32) et (33) peuvent être réunies pour composer l'équation

$$\min_u \left[\frac{\partial V(t, x)}{\partial t} + \mathcal{L}_u V(t, x) + \Lambda(t, x, u)\right] = 0. \quad (32.39)$$

La condition aux limites de l'équation (39) est la condition (34)

$$V(T, x(T)) = b(T, x(T)).$$

L'équation (39) est une *équation fonctionnelle de programmation dynamique dite équation de Bellman.*

4. Commande optimale avec information complète sur l'état du système. Système linéaire stochastique à critère quadratique de performance.' Considérons le système linéaire décrit par l'équation différentielle stochastique

$$dx = [A(t)\,x + L(t)\,u]\,dt + g(t)\,d\tilde{w}(t), \quad (32.40)$$

où, d'après (9) et (5),

$$g(t) = G(t)\,S(t), \qquad S(t)\,S^*(t) = Q(t)$$

et $Q(t)$ est une matrice symétrique définie non négative.

On demande de choisir une commande u telle qu'elle minimise la fonctionnelle (28), avec

$$\begin{aligned} b(T, x(T)) &= x^*(T) Fx(T), \\ \Lambda(t, x, u) &= x^* \Gamma(t) x + u^* N(t) u. \end{aligned} \tag{32.41}$$

Ici $\Gamma(t)$ et F sont des matrices symétriques définies non négatives, et $N(t)$, la matrice symétrique définie positive.

L'équation de Bellman (39) du problème considéré s'écrit

$$\min_u \left\{ \frac{\partial V(t, x)}{\partial t} + \frac{1}{2} \operatorname{Sp}[V_{xx} G(t) Q(t) G^*(t)] + \right.$$
$$\left. + [A(t) x + L(t) u]^* V_x + x^* \Gamma(t) x + u^* N(t) u \right\} = 0. \tag{32.42}$$

En désignant par $\mathscr{H}$ la fonction entre accolades de l'expression (42), on a

$$\frac{\partial \mathscr{H}}{\partial u_j} = \frac{\partial}{\partial u_j} \{[L(t) u]^* V_x + u^* N(t) u\} =$$
$$= \frac{\partial}{\partial u_j} \{\langle L(t) u, V_x \rangle + \langle u, N(t) u \rangle\} \tag{32.43}$$

et, donc,

$$\frac{\partial \mathscr{H}}{\partial u_j} = \sum_{k=1}^{n} L_{kj}(t) \frac{\partial V}{\partial x_k} + 2 \sum_{\mu=1}^{p} N_{j\mu}(t) u_\mu. \tag{32.44}$$

Ainsi,

$$\frac{\partial \mathscr{H}}{\partial u} = \begin{bmatrix} \partial \mathscr{H} / \partial u_1 \\ \cdots \\ \partial \mathscr{H} / \partial u_p \end{bmatrix} = L^*(t) V_x + 2N(t) u. \tag{32.45}$$

D'après (45), la condition nécessaire pour assurer l'extrémum de la fonction $\mathscr{H}(u)$

$$\frac{\partial \mathscr{H}}{\partial u} = 0 \tag{32.46}$$

devient

$$L^*(t) V_x + 2N(t) u = 0, \tag{32.47}$$

d'où

$$u = -\frac{1}{2} N^{-1}(t) L^*(t) V_x(t, x). \tag{32.48}$$

Puisque d'après (44)

$$\frac{\partial^2 \mathscr{H}}{\partial u_j \partial u_\mu} = 2N_{j\mu}(t), \tag{32.49}$$

la matrice

$$\begin{bmatrix} \frac{\partial^2 \mathcal{H}}{\partial u_1^2} & \cdots & \frac{\partial^2 \mathcal{H}}{\partial u_1 \partial u_p} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 \mathcal{H}}{\partial u_1 \partial u_p} & \cdots & \frac{\partial^2 \mathcal{H}}{\partial u_p^2} \end{bmatrix} = 2N(t), \qquad (32.50)$$

et comme nous l'avons dit plus haut, $N(t)$ est une matrice définie positive. Ainsi, la commande u définie par l'expression (48) minimise la fonction $\mathcal{H}$ et, par conséquent, c'est une commande optimale.

En remplaçant dans l'équation fonctionnelle (42) u par son expression (48) on obtient l'équation aux dérivées partielles dont la solution est recherchée sous la forme

$$V(t, x) = x^* P(t) x + p(t) \qquad (32.51)$$

avec $P(t)$, la matrice symétrique $n \times n$ et $p(t)$, la fonction scalaire.

D'après (51), on a

$$\frac{\partial V}{\partial x_j} = 2 \sum_{k=1}^{n} P_{jk}(t) x_k, \quad \frac{\partial^2 V}{\partial x_j \partial x_k} = 2P_{jk}(t) \qquad (32.52)$$

et donc,

$$V_x = 2P(t) x, \quad V_{xx} = 2P(t). \qquad (32.53)$$

L'expression (48) peut maintenant se mettre sous la forme

$$u = -N^{-1}(t) L^*(t) P(t) x. \qquad (32.54)$$

Dans ces conditions, l'équation (42) devient

$$x^* \frac{dP(t)}{dt} x + \frac{dp(t)}{dt} + \mathrm{Sp}\,[P(t) G(t) Q(t) G^*(t)] + \\ + 2[A(t) x]^* P(t) x - x^* P(t) L(t) N^{-1}(t) L^*(t) P(t) x + \\ + x^* \Gamma(t) x = 0. \qquad (32.55)$$

Etant donné que par hypothèse $P(t)$ est une matrice symétrique, on a l'identité

$$2 \langle Ax, Px \rangle = \langle x, A^* Px \rangle + \langle x, PAx \rangle. \qquad (32.56)$$

D'après (56), l'équation (55) s'écrit

$$\Big\langle x, \Big[\frac{dP(t)}{dt} + A^*(t) P(t) + P(t) A(t) - \\ - P(t) L(t) N^{-1}(t) L^*(t) P(t) + \Gamma(t)\Big] x \Big\rangle + \\ + \frac{dp(t)}{dt} + \mathrm{Sp}\,[P(t) G(t) Q(t) G^*(t)] = 0. \qquad (32.57)$$

Pour que l'équation (57) soit identiquement vérifiée par rapport à x et pour que la condition aux limites (34) soit remplie, il suffit d'observer les conditions

$$\frac{dP}{dt} = -A^*(t)P - PA(t) + PL(t)N^{-1}(t)L^*(t)P(t) - \Gamma(t) = 0, \qquad P(T) = F, \tag{32.58}$$

$$p(t) = \int_t^T \operatorname{Sp}[P(\tau)G(\tau)Q(\tau)G^*(\tau)]\,d\tau. \tag{32.59}$$

L'équation (58) est une équation différentielle matricielle de Riccati. Comme les matrices N, Γ et F intervenant dans cette équation sont symétriques, en transposant les deux membres de (58) on obtient pour la matrice P^* l'équation différentielle dont le second membre et la condition aux limites sont les mêmes que dans l'équation (58). Par suite, $P(t) = P^*(t)$, c'est-à-dire la matrice $P(t)$ est symétrique, ce que nous avons supposé dans l'expression (51).

Sous la forme explicite, la commande optimale u s'obtient en substituant à $P(t)$ dans l'expression (54) la solution de l'équation différentielle matricielle de Riccati (58).

La valeur minimale de la fonctionnelle

$$\tilde{J}(t, x) = \mathrm{M}\left\{x^*(T)Fx(T) + \int_t^T [x^*\Gamma(\tau)x + u^*N(\tau)u]\,d\tau\right\} \tag{32.60}$$

fournie par la commande optimale (54) est définie par l'expression (51) en remplaçant $P(t)$ et $p(t)$ conformément à (58) et (59).

§ 33. Commande optimale des systèmes stochastiques avec information incomplète sur l'état du système. Systèmes stochastiques linéaires à critère quadratique de performance. Théorème de séparation

Considérons le système décrit par l'équation différentielle vectorielle

$$\frac{dx}{dt} = A(t)x + L(t)u(t) + B(t)w(t). \tag{33.1}$$

Ici x est le vecteur d'état de dimension n; $A(t)$, la matrice $n \times n$; $u(t)$, le vecteur de commande de dimension p; $L(t)$, la matrice $n \times p$; $w(t)$, le processus aléatoire vectoriel de dimension r; $B(t)$, la matrice $n \times r$.

Des mesures permettent de déterminer le vecteur z de dimension m défini par la relation

$$z(t) = C(t)x(t) + v(t), \tag{33.2}$$

où $C(t)$ est la matrice $m \times n$; $v(t)$, le processus aléatoire vectoriel de dimension m.

Les processus aléatoires $w(t)$ et $v(t)$ sont des bruits blancs gaussiens à moyennes nulles. Leurs matrices de corrélation sont

$$\begin{aligned} \operatorname{cov}[w(t);\ w(\tau)] &= \mathbf{M}[w(t)\, w^*(\tau)] = Q(t)\,\delta(t-\tau), \\ \operatorname{cov}[v(t);\ v(\tau)] &= \mathbf{M}[v(t)\, v^*(\tau)] = R(t)\,\delta(t-\tau), \\ \operatorname{cov}[w(t);\ v(\tau)] &= \mathbf{M}[w(t)\, v^*(\tau)] = 0, \end{aligned} \tag{33.3}$$

où $\delta(t)$ est la fonction impulsion de Dirac; $Q(t)$, la matrice symétrique définie non négative $r \times r$; $R(t)$, la matrice symétrique définie positive $m \times m$.

Nous supposons que $x(t_0)$ soit un vecteur aléatoire gaussien de dimension n à moyenne

$$\mathbf{M}[x(t_0)] = \bar{x}_0 \tag{33.4}$$

et à matrice de corrélation

$$\mathbf{M}\{[x(t_0) - \bar{x}_0][x(t_0) - \bar{x}_0]^*\} = \Sigma_0; \tag{33.5}$$

de plus, Σ_0 est la matrice définie non négative qu'on suppose connue. En outre, on suppose que $w(t)$, $v(t)$ et $x(t_0)$ sont non corrélés.

La classe des commandes admissibles $u_j(t)$, $j = 1, \ldots, p$ est donnée ci-dessous (13). La commande $u(t)$ doit être choisie de façon à minimiser la fonctionnelle

$$J = \mathbf{M}\left\{\frac{1}{2}\langle x(T),\ S_1 x(T)\rangle + \right.$$
$$\left. + \frac{1}{2}\int_{t_0}^{T} [\langle x(s),\ Q_1(s)\, x(s)\rangle + \langle u(s),\ R_1(s)\, u(s)\rangle]\, ds\right\}. \tag{33.6}$$

Ici T est un instant *fixé*. Le but de la commande est de maintenir les moyennes des coordonnées de phase du système $x_j(t)$, $j = 1, \ldots, \ldots, n$ au voisinage du zéro, c'est-à-dire notre problème consiste à assurer la régulation de l'état du système. On suppose que S_1 et $Q_1(t)$ sont des matrices symétriques non négatives $n \times n$, et $R_1(t)$, la matrice symétrique définie positive $p \times p$.

De même qu'au § 28 désignons par $\hat{x}(t)$ l'estimation optimale au sens du critère (28.24) de l'état $x(t)$ du système (1) définie à partir des mesures (2) dans l'intervalle de temps (t_0, t). Désignons par $\tilde{x}(t)$ l'erreur d'estimation

$$\tilde{x}(t) = x(t) - \hat{x}(t). \tag{33.7}$$

Introduisons la notation $\Sigma(t)$ pour la matrice de corrélation de l'erreur d'estimation

$$\Sigma(t) = \mathbf{M}[\tilde{x}(t)\, \tilde{x}^*(t)]. \tag{33.8}$$

Théorème 1. *Pour des commandes sans prédiction, c'est-à-dire définies par les valeurs de la fonction $z(\tau)$ sur l'intervalle $t_0 \leqslant \tau \leqslant t$, l'estimation optimale $\hat{x}(t)$ de l'état $x(t)$ du système* (1) *est une solution de l'équation différentielle vectorielle*

$$\frac{d\hat{x}}{dt} = A(t)\hat{x} + L(t)u(t) + \Sigma(t)C^*(t)R^{-1}(t)[z(t) - C(t)\hat{x}],$$
$$\hat{x}(t_0) = \bar{x}_0, \qquad (33.9)$$

où $\Sigma(t)$ est la solution de l'équation différentielle matricielle de Riccati

$$\frac{d\Sigma}{dt} = A(t)\Sigma + \Sigma A^*(t) - \Sigma C^*(t)R^{-1}(t)C(t)\Sigma +$$
$$+ B(t)Q(t)B^*(t), \qquad \Sigma(t_0) = \Sigma_0. \qquad (33.10)$$

Démonstration. En plus du système (1), considérons le système auxiliaire

$$\frac{d\xi}{dt} = A(t)\xi + B(t)w(t),$$
$$\zeta(t) = C(t)\xi(t) + v(t), \qquad (33.11)$$

où

$$\xi(t_0) = x(t_0), \quad \mathbf{M}[\xi(t_0)] = \bar{x}_0,$$
$$\mathbf{M}\{[\xi(t_0) - \bar{x}_0][\xi(t_0 - \bar{x}_0]^*\} = \Sigma_0.$$

Pour le système (11) le problème de filtrage a été exploré §§ 27-29.

Les équations (1) et (11) entraînent que

$$x(t) = \xi(t) + \int_{t_0}^{t} \Phi(t, s)L(s)u(s)\,ds. \qquad (33.12)$$

Bornons-nous aux commandes $u(t)$ déterminées par les valeurs du vecteur fonction $z(\tau)$ sur l'intervalle $t_0 \leqslant \tau \leqslant t$, c'est-à-dire adoptons que

$$u(t) = \tilde{u}(t, z(\tau)), \qquad \tau \in [t_0, t]. \qquad (33.13)$$

Dans ce cas bien que d'après (2) la commande $u(t)$ constitue un processus aléatoire, la valeur concrète de sa *réalisation* à l'instant t est connue. C'est pourquoi on a la relation

$$\hat{x}(t) = \hat{\xi}(t) + \int_{t_0}^{t} \Phi(t, s)L(s)u(s)\,ds. \qquad (33.14)$$

La fonction $\hat{\xi}(t)$ étant solution du problème de filtrage optimal étudié aux §§ 27-29, elle vérifie l'équation différentielle

$$\frac{d\hat{\xi}}{dt} = A(t)\hat{\xi} + \Sigma(t)C^*(t)R^{-1}(t)[\zeta(t) - C(t)\hat{\xi}], \qquad \hat{\xi}(t_0) = \bar{x}_0, \quad (33.15)$$

où $\Sigma(t)$ est solution de l'équation différentielle matricielle de Riccati (10).

En dérivant par rapport à t les deux membres de l'équation (14) et en retenant que

$$\frac{d}{dt}\Phi(t, s) = \frac{d\theta(t)}{dt}\theta^{-1}(s) = A(t)\theta(t)\theta^{-1}(s) = A(t)\Phi(t, s),$$

$$\Phi(t, t) = E, \quad \frac{d}{dt}\int_{t_0}^{t}\Phi(t, s)L(s)u(s)\,ds =$$

$$= A(t)\int_{t_0}^{t}\Phi(t, s)L(s)u(s)\,ds + \Phi(t, t)L(t)u(t) =$$

$$= A(t)[\hat{x}(t) - \hat{\xi}(t)] + L(t)u(t), \qquad (33.16)$$

on a

$$\frac{d\hat{x}}{dt} = \frac{d\hat{\xi}}{dt} + A(t)[\hat{x}(t) - \hat{\xi}(t)] + L(t)u(t),$$

ou, d'après (15),

$$\frac{d\hat{x}}{dt} = A(t)\hat{x} + L(t)u(t) + \Sigma(t)C^*(t)R^{-1}(t)[\zeta(t) - C(t)\hat{\xi}],$$

$$\hat{\xi}(t_0) = \bar{x}_0, \qquad (33.17)$$

avec $\Sigma(t)$, solution de l'équation de Riccati (10).

(13) et (14) entraînent que

$$x - \hat{x} = \xi - \hat{\xi}. \qquad (33.18)$$

Donc,

$$\zeta(t) - C(t)\hat{\xi} = C(t)(\xi - \hat{\xi}) + v(t) =$$

$$= C(t)(x - \hat{x}) + v(t) = z(t) - C(t)\hat{x}. \qquad (33.19)$$

Constatons encore que d'après (14),

$$\hat{x}(t_0) = \hat{\xi}(t_0). \qquad (33.20)$$

Ainsi, l'équation (17) se met sous la forme (9)

$$\frac{d\hat{x}}{dt} = A(t)\hat{x} + L(t)u(t) + \Sigma(t)C^*(t)R^{-1}(t)[z(t) - C(t)\hat{x}],$$

$$\hat{x}(t_0) = \bar{x}_0,$$

où $\Sigma(t)$ est solution de l'équation différentielle matricielle de Riccati (10). Le théorème est démontré.

Lemme 1. *La fonctionnelle*

$$J(x, t) = \mathbf{M}\left\{\frac{1}{2}\langle x(T), S_1 x(T)\rangle + \right.$$

$$\left. + \frac{1}{2}\int_{t}^{T}[\langle x(s), Q_1(s)x(s)\rangle + \langle u(s), R_1(s)u(s)\rangle]\,ds\right\} \qquad (33.21)$$

peut être mise sous la forme

$$J(x,\ t)=\mathbf{M}\left\{\frac{1}{2}\langle\hat{x}(T),\ \ S_1\hat{x}(T)\rangle+\right.$$

$$\left.+\frac{1}{2}\int_t^T[\langle\hat{x}(s),\ Q_1(s)\hat{x}(s)\rangle+\langle u(s),\ R_1(s)u(s)\rangle]\,ds\right\}+$$

$$+\frac{1}{2}\operatorname{Sp}[S_1\Sigma(T;\ \Sigma_0,\ t_0)]+\frac{1}{2}\int_t^T\operatorname{Sp}[Q_1(s)\Sigma(s;\ \Sigma_0,\ t_0])\,ds. \qquad (33.22)$$

D é m o n s t r a t i o n. D'après (7),

$$x(t)=\hat{x}(t)+\tilde{x}(t).$$

Donc,

$$\mathbf{M}[\langle x(s),\ Q_1(s)x(s)\rangle]=$$
$$=\mathbf{M}\{[\hat{x}^*(s)+\tilde{x}^*(s)][Q_1(s)\hat{x}(s)+Q_1(s)\tilde{x}(s)]\}=$$
$$=\mathbf{M}[\hat{x}^*(s)Q_1(s)\hat{x}(s)]+\mathbf{M}[\tilde{x}^*(s)Q_1(s)\tilde{x}(s)]+$$
$$+\mathbf{M}[\tilde{x}^*(s)Q_1(s)\hat{x}(s)]+\mathbf{M}[\hat{x}^*(s)Q_1(s)\tilde{x}(s)]. \qquad (33.23)$$

Remarquons que puisque Q_1 est une matrice symétrique, on a l'identité

$$\tilde{x}^*Q_1\hat{x}=(Q_1\tilde{x})^*\hat{x}=\langle Q_1\tilde{x},\ \hat{x}\rangle=\langle\hat{x},\ Q_1\tilde{x}\rangle=\hat{x}^*Q_1\tilde{x}.$$

Comme

$$\tilde{x}^*(s)\hat{x}(s)=\operatorname{Sp}[\hat{x}(s)\tilde{x}^*(s)],\quad \mathbf{M}[\tilde{x}^*(s)Q_1(s)\hat{x}(s)]=$$
$$=\mathbf{M}\{\operatorname{Sp}[Q_1(s)\hat{x}(s)\tilde{x}^*(s)]\}=\operatorname{Sp}\{Q_1(s)\mathbf{M}[\hat{x}(s)\tilde{x}^*(s)]\} \qquad (33.24)$$

et, d'après (27.35),

$$\mathbf{M}[\hat{x}(s)\tilde{x}^*(s)]=0,$$

il vient

$$\mathbf{M}[\tilde{x}^*(s)Q_1(s)\hat{x}(s)]=\mathbf{M}[\hat{x}^*(s)Q_1(s)\tilde{x}(s)]=0,$$

et la relation (23) s'écrit

$$\mathbf{M}[\langle x(s),\ Q_1(s)x(s)\rangle]=\mathbf{M}[\hat{x}^*(s)Q_1(s)\hat{x}(s)]+$$
$$+\mathbf{M}[\tilde{x}^*(s)Q_1(s)\tilde{x}(s)]. \qquad (33.25)$$

D'une façon analogue à (24), on a la relation

$$\mathbf{M}[\tilde{x}^*(s)Q_1(s)\tilde{x}(s)]=\operatorname{Sp}\{Q_1(s)\mathbf{M}[\tilde{x}(s)\tilde{x}^*(s)]\},$$

qui, conformément à (8) et (10), devient

$$\mathbf{M}[\tilde{x}^*(s)Q_1(s)\tilde{x}(s)]=\operatorname{Sp}[Q_1(s)\Sigma(s;\ \Sigma_0,\ t_0). \qquad (33.26)$$

Ainsi, en vertu de (25) et (26), on a

$$M[\langle x(s), Q_1(s)x(s)\rangle] = M[\langle \hat{x}(s), Q_1(s)\hat{x}(s)\rangle] + \\ + \mathrm{Sp}[Q_1(s)\Sigma(s;\Sigma_0, t_0)]. \quad (33.27)$$

D'une façon analogue à (27),

$$M[\langle x(T), S_1x(T)\rangle] = M[\langle \hat{x}(T), S_1\hat{x}(T)\rangle] + \\ + \mathrm{Sp}[S_1\Sigma(T;\Sigma_0, t_0)]. \quad (33.28)$$

Les relations (27) et (28) justifient l'expression (22). Le lemme est démontré.

Théorème 2 (théorème de séparation). *Dans le système décrit par les équations* (1) *et* (2)

$$\frac{dx}{dt} = A(t)x + L(t)u(t) + B(t)w(t),$$
$$z(t) = C(t)x(t) + v(t),$$

la commande optimale $u^0(t)$ *minimisant la fonctionnelle* (6)

$$J = M\left\{\frac{1}{2}\langle x(T), S_1x(T)\rangle + \\ + \frac{1}{2}\int_{t_0}^{T}[\langle x(s), Q_1(s)x(s)\rangle + \langle u(s), R_1(s)u(s)\rangle]\,ds\right\}$$

est de la forme

$$u^0(t) = -R_1^{-1}(t)L^*(t)K(t)\hat{x}(t;\bar{x}_0, t_0), \quad (33.29)$$

où $K(t)$ *est solution de l'équation différentielle matricielle de Riccati*

$$\frac{dK}{dt} = -KA(t) - A^*(t)K + KL(t)R_1^{-1}(t)L^*(t)K - Q_1(t), \\ K(T) = S_1. \quad (33.30)$$

Ici $\hat{x}(t;\bar{x}_0, t_0)$ *est solution de l'équation différentielle vectorielle*

$$\frac{d\hat{x}}{dt} = A(t)\hat{x} + L(t)u^0(t) + \Sigma(t)C^*(t)R^{-1}(t)[z(t) - C(t)\hat{x}], \\ \hat{x}(t_0) = \bar{x}_0 \quad (33.31)$$

et $\Sigma(t)$, *solution de l'équation différentielle matricielle de Riccati*

$$\frac{d\Sigma}{dt} = A(t)\Sigma + \Sigma A^*(t) - \Sigma C^*(t)R^{-1}(t)C(t)\Sigma + \\ + B(t)Q(t)B^*(t), \quad \Sigma(t_0) = \Sigma_0. \quad (33.32)$$

Dans ces conditions la valeur minimale de la fonctionnelle

$$J(x, t) = M\left\{\frac{1}{2}\langle x(T), S_1x(T)\rangle + \\ + \frac{1}{2}\int_{t}^{T}[\langle x(s), Q_1(s)x(s)\rangle + \langle u(s), R_1(s)u(s)\rangle]\,ds\right\} \quad (33.33)$$

ou de la fonctionnelle équivalente (22) *est*

$$\tilde{J}(\hat{x}, t) = \frac{1}{2}\left\{\langle \hat{x}(t), K(t)\hat{x}(t)\rangle + \int_t^T \operatorname{Sp}[\Sigma(s) C^*(s) R^{-1}(s) C(s) \Sigma(s) K(s)]\, ds + \operatorname{Sp}[S_1 \Sigma(T; \Sigma_0, t_0)] + \int_t^T \operatorname{Sp}[Q_1(s)\Sigma(s; \Sigma_0, t_0)]\, ds\right\}. \quad (33.34)$$

D é m o n s t r a t i o n. En tenant compte de l'expression (22) de la fonctionnelle minimisée, l'équation de Bellman peut se mettre sous la forme:

$$-\frac{\partial \tilde{J}}{\partial t} = \min_u \left\{\frac{1}{2}\langle \hat{x}(t), Q_1(t)\hat{x}(t)\rangle + \frac{1}{2}\operatorname{Sp}[Q_1(t)\Sigma(t; \Sigma_0, t_0)] + \frac{1}{2}\langle u, R_1(t) u\rangle + \langle A(t)\hat{x}(t), \operatorname{grad} \tilde{J}(\hat{x}, t)\rangle + \langle L(t) u, \operatorname{grad} \tilde{J}(\hat{x}, t)\rangle + \frac{1}{2}\operatorname{Sp}[\Sigma(t) C^*(t) R^{-1}(t) C(t)\Sigma(t) \mathcal{N}(t)]\right\}, \quad (33.35)$$

où

$$\mathcal{N}(t) = \begin{bmatrix} \frac{\partial^2 \tilde{J}}{\partial \hat{x}_1^2} & \frac{\partial^2 \tilde{J}}{\partial \hat{x}_1 \partial \hat{x}_2} & \cdots & \frac{\partial^2 \tilde{J}}{\partial \hat{x}_1 \partial \hat{x}_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \tilde{J}}{\partial \hat{x}_n \partial \hat{x}_1} & \frac{\partial^2 \tilde{J}}{\partial \hat{x}_n \partial \hat{x}_2} & \cdots & \frac{\partial^2 \tilde{J}}{\partial \hat{x}_n^2} \end{bmatrix}. \quad (33.36)$$

D'après (22) et (30), la condition aux limites de l'équation (35) est

$$\tilde{J}(\hat{x}(T), T) = \frac{1}{2}\langle x(T), S_1 x(T)\rangle = \frac{1}{2}\langle \hat{x}(T), K(T)\hat{x}(T)\rangle + \frac{1}{2}\operatorname{Sp}[\Sigma(T; \Sigma_0, t_0) K(T)]. \quad (33.37)$$

L'équation de Bellman (35) peut s'obtenir de la façon suivante. Dans le problème en question on demande de minimiser la fonctionnelle (22); on entend par état du système le vecteur fonction $\hat{x}(t)$ défini par l'équation différentielle stochastique (9).

En désignant d'après (28.10) et (28.96)

$$\chi(t) = z(t) - C(t) x(t)$$

et en tenant compte de ce que d'après (28.79)

$$G(t) = \Sigma(t) C^*(t) R^{-1}(t)$$

on peut récrire l'équation (9) sous la forme

$$\frac{d\hat{x}}{dt} = A(t)\hat{x} + L(t) u(t) + G(t)\chi(t), \quad \hat{x}(t_0) = \bar{x}_0.$$

Nous avons montré précédemment (28.112) que le processus aléatoire $\chi(t)$ est un bruit blanc gaussien à moyenne nulle et à matrice de corrélation

$$\operatorname{cov}[\chi(t);\ \chi(\tau)] = M[\chi(t)\,\chi^*(\tau)] = R(t)\,\delta(t-\tau),$$

où $R(t)$ est la matrice symétrique définie positive qui fait partie de l'expression (3).

Conformément à (32.21), l'opérateur différentiel $\hat{\mathscr{L}}_u$ du processus aléatoire markovien $\hat{x}(t)$ se définit à l'aide de la relation

$$\hat{\mathscr{L}}_u\tilde{J}(t,\ \hat{x}) = \frac{1}{2}\operatorname{Sp}[\tilde{J}_{\hat{x}\hat{x}}(t,\ \hat{x})\,G(t)\,R(t)\,G^*(t)] + \\ + \langle [A(t)\,\hat{x} + L(t)\,u],\ \operatorname{grad}\tilde{J}(\hat{x},\ t)\rangle.$$

L'équation de Bellman du problème de minimisation de la fonctionnelle (22) est de la forme (32.39)

$$\min_u\left[\frac{\partial\tilde{J}(t,\ \hat{x})}{\partial t} + \hat{\mathscr{L}}_u\tilde{J}(t,\ \hat{x}) + \hat{\Lambda}(t,\ \hat{x},\ u)\right] = 0,$$

où $\hat{\Lambda}(t,\ \hat{x},\ u)$ est la fonction sous le signe somme de la fonctionnelle (22):

$$\hat{\Lambda}(t,\ \hat{x}, u) = \frac{1}{2}\langle\hat{x}(t),\ Q_1(t)\,\hat{x}(t)\rangle + \frac{1}{2}\langle u,\ R_1(t)\,u\rangle + \frac{1}{2}\operatorname{Sp}[Q_1(t)\,\Sigma(t;\Sigma_0,\ t_0)].$$

D'après (36),

$$\tilde{J}_{\hat{x}\hat{x}}(t,\ \hat{x}) = \mathscr{N}(t).$$

Comme $\mathscr{N}(t)$ est une matrice symétrique, on a la relation

$$\operatorname{Sp}[\tilde{J}_{\hat{x}\hat{x}}(t,\ \hat{x})\,G(t)\,R(t)\,G^*(t)] = \\ = \operatorname{Sp}[\mathscr{N}(t)\,\Sigma(t)\,C^*(t)\,R^{-1}(t)\,R(t)\,R^{-1}(t)\,C(t)\,\Sigma(t)] = \\ = \operatorname{Sp}[\Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma(t)\,\mathscr{N}(t)].$$

Ainsi, dans le problème considéré, l'équation de Bellman se met sous la forme (35).

En désignant

$$\mathscr{H} = \frac{1}{2}\langle\hat{x}(t),\ Q_1(t)\,\hat{x}(t)\rangle + \frac{1}{2}\operatorname{Sp}[Q_1(t)\,\Sigma(t;\ \Sigma_0,\ t_0)] + \\ + \frac{1}{2}\langle u,\ R_1(t)\,u\rangle + \langle A(t)\,\hat{x}(t),\ \operatorname{grad}\tilde{J}(\hat{x},\ t)\rangle + \\ + \langle L(t)\,u,\ \operatorname{grad}\tilde{J}(\hat{x},\ t)\rangle + \\ + \frac{1}{2}\operatorname{Sp}[\Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma(t)\,\mathscr{N}(t)], \qquad (33.38)$$

on trouve la valeur de u qui minimise la fonction $\mathscr{H}$. Puisque

$$\frac{\partial\mathscr{H}}{\partial u_j} = \sum_{l=1}^{p}(R_1)_{jl}\,u_l + \sum_{k=1}^{n}L_{kj}\frac{\partial\tilde{J}}{\partial\hat{x}_k}, \qquad j = 1,\ \ldots,\ p, \qquad (33.39)$$

les conditions nécessaires de l'extrémum

$$\frac{\partial\mathscr{H}}{\partial u_j} = 0, \qquad j = 1,\ \ldots,\ p \qquad (33.40)$$

deviennent

$$\sum_{l=1}^{p} (R_1)_{jl} u_l = -\sum_{k=1}^{n} L_{kj} \frac{\partial \tilde{J}}{\partial \hat{x}_k}, \qquad j = 1, \ldots, p. \qquad (33.41)$$

Le système d'équations scalaires (41) est équivalent à l'équation vectorielle

$$R_1(t)\, u(t) = -L^*(t) \operatorname{grad} \tilde{J}(\hat{x}, t). \qquad (33.42)$$

Comme $R_1(t)$ est la matrice définie positive, l'équation (42) implique que la commande $u(t)$ extrémisant la fonction $\mathcal{H}$ soit de la forme

$$u(t) = -R_1^{-1}(t)\, L^*(t) \operatorname{grad} \tilde{J}(\hat{x}, t). \qquad (33.43)$$

D'après (39)

$$\frac{\partial^2 \mathcal{H}}{\partial u_j\, \partial u_k} = (R_1)_{jk}, \qquad j, k = 1, \ldots, r \qquad (33.44)$$

et, par conséquent, la matrice $n \times n$ dont les éléments sont $\partial^2 \mathcal{H} / \partial u_j \partial u_k$, $j, k = 1, \ldots, r$ est définie positive. Ainsi, la commande $u(t)$ déterminée par l'expression (43) minimise la fonction $\mathcal{H}$.

Montrons maintenant que la solution de l'équation de Bellman (35) s'écrit (34):

$$\tilde{J}(\hat{x}, t) = \frac{1}{2} \langle \hat{x}(t), K(t)\, \hat{x}(t) \rangle + \\ + \frac{1}{2} \int_t^T \operatorname{Sp}[\Sigma(s)\, C^*(s)\, R^{-1}(s)\, C(s)\, \Sigma(s)\, K(s)]\, ds + \\ + \frac{1}{2} \operatorname{Sp}[\Sigma(T; \Sigma_0, t_0)\, K(T)] + \frac{1}{2} \int_t^T \operatorname{Sp}[Q_1(s)\, \Sigma(s; \Sigma_0, t_0)]\, ds$$

où $K(t)$ est la matrice $n \times n$, solution de l'équation de Riccati (30).

L'expression (34) conduit à

$$\operatorname{grad} \tilde{J}(\hat{x}, t) = K(t)\, \hat{x}(t). \qquad (33.45)$$

En outre, la commande optimale $u^0(t)$ est d'après (43)

$$u^0(t) = -R_1^{-1}(t)\, L^*(t)\, K(t)\, \hat{x}(t). \qquad (33.46)$$

En vertu de (34),

$$\frac{\partial^2 \tilde{J}}{\partial \hat{x}_j\, \partial \hat{x}_k} = K_{jk}(t) \qquad (33.47)$$

et, par suite,

$$\mathcal{N}(t) = K(t). \qquad (33.48)$$

Il résulte de (34) que

$$\frac{\partial \tilde{J}(\hat{x}, t)}{\partial t} = \frac{1}{2}\hat{x}^* \frac{dK}{dt}\hat{x} -$$
$$- \frac{1}{2}\,\mathrm{Sp}\,[\Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma(t)\,K(t)] -$$
$$- \frac{1}{2}\,\mathrm{Sp}\,[Q_1(t)\,\Sigma(t;\ \Sigma_0,\ t_0)]. \quad (33.49)$$

L'expression (30) donne

$$\frac{1}{2}\hat{x}^* \frac{dK}{dt}\hat{x} = \frac{1}{2}\hat{x}^*(-KA - A^*K + KLR_1^{-1}L^*K - Q_1)\,\hat{x} =$$
$$= -\frac{1}{2}\hat{x}^*KA\hat{x} - \frac{1}{2}\hat{x}^*A^*K\hat{x} +$$
$$+ \frac{1}{2}\hat{x}^*KLR_1^{-1}L^*K\hat{x} - \frac{1}{2}\hat{x}^*Q_1\hat{x}. \quad (33.50)$$

Comme K est une matrice symétrique, l'expression (50) devient

$$\frac{1}{2}\hat{x}^* \frac{dK}{dt}\hat{x} = -\langle A\hat{x},\ K\hat{x}\rangle + \frac{1}{2}\hat{x}^*KLR_1^{-1}L^*K\hat{x} - \frac{1}{2}\hat{x}^*Q_1\hat{x}. \quad (33.51)$$

Il résulte de (51) et (49) que

$$-\frac{\partial \tilde{J}(\hat{x}, t)}{\partial t} = \langle A\hat{x},\ K\hat{x}\rangle - \frac{1}{2}\hat{x}^*K(t)\,L(t)\,R_1^{-1}(t)\,L^*(t)\,K(t)\,\hat{x} +$$
$$+ \frac{1}{2}\,\mathrm{Sp}\,[\Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma(t)\,K(t)] +$$
$$+ \frac{1}{2}\langle\hat{x},\ Q_1(t)\,\hat{x}\rangle + \frac{1}{2}\,\mathrm{Sp}\,[Q_1(t)\,\Sigma(t;\ \Sigma_0,\ t_0)]. \quad (33.52)$$

Si l'on porte d'après (46) dans le second membre de l'équation de Bellman (35) la valeur optimale $u^0(t)$

$$u^0(t) = -R_1^{-1}(t)\,L^*(t)\,K(t)\,\hat{x},$$

on obtient l'expression qui représente $\min\limits_u \mathscr{H}$, où $\mathscr{H}$ est de la forme (4). Cette expression s'écrit

$$\min_u \mathscr{H} = \frac{1}{2}\langle\hat{x},\ Q_1(t)\,\hat{x}\rangle + \frac{1}{2}\,\mathrm{Sp}\,[Q_1(t)\,\Sigma(t;\ \Sigma_0,\ t_0)] +$$
$$+ \langle A\hat{x},\ K\hat{x}\rangle - \frac{1}{2}\hat{x}^*K(t)\,L(t)\,R_1^{-1}(t)\,L^*(t)\,K(t)\,\hat{x} +$$
$$+ \frac{1}{2}\,\mathrm{Sp}\,[\Sigma(t)\,C^*(t)\,R^{-1}(t)\,C(t)\,\Sigma(t)\,K(t)]. \quad (33.53)$$

Les expressions (52) et (53) coïncident, et par suite, l'équation de Bellman (35) est vérifiée, c'est-à-dire l'expression (34) est en effet solution de l'équation de Bellman (35) Le théorème est démontré.

§ 34. Commande optimale des systèmes stochastiques discontinus

1. Systèmes linéaires à critère quadratique de performance avec information complète sur l'état du système. Considérons le système décrit par l'équation vectorielle aux différences finies

$$x(t_{n+1}) = \Phi(t_{n+1}, t_n)\, x(t_n) + L(t_n)\, u_n + \Gamma(t_n)\, w_n,$$
$$n = 0, 1, 2, \ldots, \qquad x(t_0) = s. \tag{34.1}$$

Ici x est le vecteur d'état de dimension n; $\Phi(t_{n+1}, t_n)$, la matrice $n \times n$; $L(t_n)$, la matrice $n \times p$; u_n, le vecteur de commande de dimension p; $\Gamma(t_n)$, la matrice $n \times r$; w_n, la suite vectorielle de dimension r des variables aléatoires gaussiennes réciproquement indépendantes à moyenne nulle.

La matrice de corrélation de la suite discontinue w_n s'écrit

$$\mathrm{M}\,[w_n w_k^*] = Q_n \delta_{nk}, \tag{34.2}$$

où Q_n est la matrice symétrique définie non négative $r \times r$, et δ_{nk}, le symbole de Kronecker

$$\delta_{nk} = \begin{cases} 1 \text{ pour } n = k; \\ 0 \text{ pour } n \neq k. \end{cases}$$

Désigné par s, le vecteur aléatoire gaussien de dimension n à moyenne

$$\mathrm{M}\,[s] = \mu \tag{34.3}$$

et à matrice de corrélation

$$\mathrm{M}\,[(s - \mu)(s - \mu)^*] = P(t_0). \tag{34.4}$$

Ici $P(t_0)$ est la matrice définie non négative $n \times n$ supposée connue. On suppose de plus que $w(t)$ et s sont non corrélés.

La commande u_n $(n = 0, 1, \ldots, N - 1)$ doit être choisie de façon à minimiser la fonctionnelle

$$J = \mathrm{M}\,\Big\{\langle x(t_N), \mathscr{F} x(t_N)\rangle + \sum_{n=0}^{N-1} [\langle x(t_n), \mathfrak{M}_n x(t_n)\rangle + \langle u_n, \mathfrak{N}_n u_n\rangle]\Big\}. \tag{34.5}$$

Ici N est fixé, c'est-à-dire t_N est un instant fixé. Le but de la commande est de maintenir les valeurs moyennes des coordonnées de phase du système $x_j(t_n)$ près du zéro, c'est-à-dire le problème consiste à assurer la régulation de l'état du système. On suppose que $\mathscr{F}$ et $\mathfrak{M}_n$ sont des matrices symétriques définies non négatives $n \times n$, et $\mathfrak{N}_n$ la matrice symétrique définie positive $p \times p$.

Théorème 1. *Pour les commandes admissibles u_n, fonctions de $x(t_n)$ et t_n, la commande optimale u_n^0, c'est-à-dire la commande*

minimisant la fonctionnelle (5), *est*

$$u_n^0 = -\mathscr{B}(t_n)\, x(t_n), \tag{34.6}$$

où la matrice $\mathscr{B}(t_n)$ *du type* $p \times n$ *est de la forme*

$$\mathscr{B}(t_n) = [\mathfrak{N}_n + L^*(t_n)\, \mathscr{S}(t_{n+1})\, L(t_n)]^{-1} L^*(t_n)\, \mathscr{S}(t_{n+1})\, \Phi(t_{n+1}, t_n). \tag{34.7}$$

La matrice $\mathscr{S}(t_n)$ *de type* $n \times n$ *faisant partie de l'expression* (7) *est définie par la relation récurrente*

$$\begin{aligned}\mathscr{S}(t_n) &= \Phi^*(t_{n+1}, t_n)\, \mathscr{S}(t_{n+1})\, \Phi(t_{n+1}, t_n) + \mathfrak{M}_n - \\ &\quad - \mathscr{B}^*(t_n)\, [\mathfrak{N}_n + L^*(t_n)\, \mathscr{S}(t_{n+1})\, L(t_n)]\, \mathscr{B}(t_n) = \\ &= [\Phi(t_{n+1}, t_n) - L(t_n)\, \mathscr{B}(t_n)]^*\, \mathscr{S}(t_{n+1})\, \Phi(t_{n+1}, t_n) + \mathfrak{M}_n = \\ &= [\Phi(t_{n+1}, t_n) - L(t_n)\, \mathscr{B}[(t_n)]^*\, \mathscr{S}(t_{n+1})\, [\Phi(t_{n+1}, t_n) \\ &\quad - L(t_n)\, \mathscr{B}(t_n)] + \mathscr{B}^*(t_n)\, \mathfrak{N}_n \mathscr{B}(t_n) + \mathfrak{M}_n \end{aligned} \tag{34.8}$$

et par la condition à l'instant t_N

$$\mathscr{S}(t_N) = \mathscr{F}. \tag{34.9}$$

La valeur minimale de la fonctionnelle (5) *s'écrit*

$$\min J = \mu^* \mathscr{S}(t_0)\, \mu + \operatorname{Sp} [\mathscr{S}(t_0)\, P(t_0)] + \sum_{n=0}^{N-1} \operatorname{Sp} [\Gamma^*(t_n)\, \mathscr{S}(t_{n+1})\, \Gamma(t_n)\, Q_n]. \tag{34.10}$$

Quelques remarques préliminaires. Avant de passer à la démonstration du théorème 1, examinons quelques problèmes auxiliaires.

1°. Soient $x \in X$ et $y \in Y$ des variables aléatoires scalaires sur un espace de probabilité.

Adoptons comme commandes admissibles u les commandes appartenant à l'ensemble U

$$u \in U.$$

Soit une fonction scalaire positive

$$l = l(x, y, u).$$

Adoptons comme critère de performance

$$\mathrm{M}l(x, y, u),$$

l'espérance mathématique d'après la distribution commune des variables aléatoires x et y.

Supposons que

$$\min_{u(x, y)} \mathrm{M}l(x, y, u)$$

est le minimum de $\mathrm{M}l(x, y, u)$ sur l'ensemble des commandes admissibles $u \in U$. Alors on a le lemme suivant.

Lemme 1. *Supposons que la fonction scalaire* $l(x, y, u)$ *possède un minimum unique sur l'ensemble des* $u \in U$ *pour tout* $x \in X$ *et tout* $y \in Y$. *Soit* $u^0(x, y)$ *la valeur de* u *qui réalise ce minimum. Alors*

$$\min_{u(x, y)} \mathrm{M}l(x, y, u) = \mathrm{M}l(x, y, u^0(x, y)) = \mathrm{M} \min_u l(x, y, u). \tag{34.11}$$

D é m o n s t r a t i o n d u l e m m e. Par condition, pour tout $u \in U$ on a

$$l(x, y, u) \geqslant l(x, y, u^0(x, y)) = \min_u l(x, y, u). \tag{34.12}$$

Par conséquent,

$$\mathrm{M}l(x, y, u) \geqslant \mathrm{M}l(x, y, u^0(x, y)) = \mathrm{M} \min_u l(x, y, u). \tag{34.13}$$

En minimisant par rapport à u (où $u \in U$) le premier membre de l'inégalité (13), on a

$$\min_{u(x, y)} \mathrm{M}l(x, y, u) \geqslant \mathrm{M}l(x, y, u^0(x, y)) = \mathrm{M} \min_u l(x, y, u). \tag{34.14}$$

Puisque $u^0(x, y) \in U$, on obtient l'inégalité

$$\mathrm{M}l(x, y, u^0(x, y)) \geqslant \min_{u(x, y)} \mathrm{M}l(x, y, u). \tag{34.15}$$

La comparaison des inégalités (14) et (15) conduit à la relation (11);

$$\min_{u(x, y)} \mathrm{M}l(x, y, u) = \mathrm{M}l(x, y, u^0(x, y)) = \mathrm{M} \min_u l(x, y, u).$$

Le lemme est démontré

Constatons que l'opération $\min_u l(x, y, u)$ détermine u comme une fonction de x et y et que le lemme démontré affirme que la minimisation sur l'ensemble U et le calcul de l'espérance mathématique d'après la distribution commune des variables aléatoires x et y sont commutatifs

2°. Supposons, de même que précédemment, que $x \in X$ et $y \in Y$ sont des variables aléatoires scalaires définies dans un espace de probabilité. Pourtant admettons cette fois que le choix de la commande soit basé sur l'information relative à la seule variable y. Adoptons comme commandes admissibles les commandes $u = u(y)$ qui appartiennent à l'ensemble U donné

$$u = u(y) \in U$$

De même que dans ce qui précède, $l = l(x, y, u)$ est une fonction scalaire positive donnée

Soit $\min_{u(y)} \mathrm{M}l(x, y, u)$ le minimum de $\mathrm{M}l(x, y, u)$ sur l'ensemble des commandes admissibles $u = u(y) \in U$. Alors on a le lemme suivant.

L e m m e 2. *Soit* $\mathrm{M}[\cdot \mid y]$ *l'espérance mathématique conditionnelle de y donné. Supposons que la fonction scalaire* $f(y, u) = \mathrm{M}[l(x, y, u) \mid y]$ *possède un minimum unique sur l'ensemble* $u \in U$ *pour tout* $y \in Y$. *Soit* $u^0(y)$ *la valeur de u qui assure ce minimum. Alors*

$$\min_{u(y)} l(x, y, u) = \mathrm{M}l(x, y, u^0(y)) =$$
$$= \mathrm{M}_y \{\min \mathrm{M}[l(x, y, u) \mid y]\}, \tag{34.16}$$

où le symbole M_y *désigne l'espérance mathématique d'après la distribution de la variable aléatoire y.*

D é m o n s t r a t i o n d u l e m m e. Par condition on a pour tout $u \in U$

$$f(y, u) \geqslant f(y, u^0(y)) = \min_u f(y, u). \tag{34.17}$$

Par suite,

$$\mathrm{M}l(x, y, u) = \mathrm{M}_y f(y, u) = \mathrm{M}_y f(y, u^0(y)) =$$
$$= \mathrm{M}l(x, y, u^0(y)) = \mathrm{M}_y \{\min_u \mathrm{M}[l(x, y, u) \mid y]\}. \tag{34.18}$$

En minimisant par rapport à u, où $u \in U$, le premier membre de l'inégalité (18), on obtient

$$\min_{u(y)} \mathrm{M}l(x, y, u) \geqslant \mathrm{M}l(x, y, u^0(y)) =$$
$$= \mathrm{M}_y \{\min_u \mathrm{M}[l(x, y, u) \mid y]\}. \tag{34.19}$$

Puisque $u^0(y) \in U$, on a l'inégalité

$$\mathrm{M}l(x, y, u^0(y)) \geqslant \min_{u(y)} \mathrm{M}l(x, y, u). \tag{34.20}$$

La comparaison des inégalités (19) et (20) conduit à la relation (15)

$$\min_{u(y)} \mathrm{M}l(x, y, u) = \mathrm{M}l(x, y, u^0(y)) = \mathrm{M}_y \{\min_u \mathrm{M}[l(x, y, u) \mid y]\}.$$

Le lemme est démontré.

Remarquons que l'opération $\min_u [l(x, y, u) \mid y]$ détermine u comme fonction de y. Le lemme démontré affirme que le calcul de l'espérance mathématique d'après la distribution de la variable aléatoire y et la minimisation sur l'ensemble de $u = u(y) \in U$ sont des opérations commutatives.

3°. Dans ce qui suit nous aurons besoin de la moyenne de la forme quadratique des variables aléatoires gaussiennes. A cet effet, on a le lemme suivant.

L e m m e 3. *Soit x le vecteur aléatoire de dimension n et soit*

$$\mathrm{M}[x] = m, \qquad \mathrm{M}[(x - m)(x - m)^*] = R. \tag{34.21}$$

Alors, pour toute matrice symétrique S de type $n \times n$ on a la relation

$$\mathrm{M}[x^*Sx] = m^*Sm + \mathrm{Sp}[SR]. \tag{34.22}$$

D é m o n s t r a t i o n. Puisque

$$\mathrm{M}[x^*Sx] = \mathrm{M}[(x - m)^* S (x - m)] + \mathrm{M}[x^*Sm] + \mathrm{M}[m^*Sx] - m^*Sm,$$

si l'on tient compte de $\mathrm{M}[x] = m$, on obtient la relation

$$\mathrm{M}[x^*Sx] = \mathrm{M}[(x - m)^* S (x - m)] + m^*Sm. \tag{34.23}$$

Quels que soient les vecteurs a et b de dimension n et la matrice symétrique S de type $n \times n$ on a l'identité

$$a^*Sb = \mathrm{Sp}[Sab^*],$$

d'où l'on tire que

$$(x - m)^* S (x - m) = \mathrm{Sp}[S (x - m)(x - m)^*]. \tag{34.24}$$

Conformément à (24)

$$\mathrm{M}[(x - m)^* S (x - m)] = \mathrm{Sp}\{\mathrm{M}[S (x - m)(x - m)^*]\} =$$
$$= \mathrm{Sp}\{S[\mathrm{M}(x - m)(x - m)^*]\} = \mathrm{Sp}[SR]. \tag{34.25}$$

Les relations (25) et (23) entraînent la proposition (22)

$$\mathrm{M}[x^*Sx] = m^*Sm + \mathrm{Sp}[SR].$$

Le lemme est démontré.

D é m o n s t r a t i o n d u t h é o r è m e 1. En éliminant des relations (7) et (8) la matrice $\mathscr{R}(t_n)$ de type $p \times n$ on obtient que la matrice symétrique $\mathscr{S}(t_n)$ du type $n \times n$ vérifie l'équation

aux différences

$$\mathcal{S}(t_n) = \Phi^*(t_{n+1}, t_n)\,\mathcal{S}(t_{n+1})\,\Phi(t_{n+1}, t_n) + \mathfrak{M}_n - \\ - \Phi^*(t_{n+1}, t_n)\,\mathcal{S}(t_{n+1})\,L(t_n)\,[\mathfrak{N}_n + L^*(t_n)\,\mathcal{S}(t_{n+1})\,L(t_n)]^{-1} \times \\ \times L^*(t_n)\,\mathcal{S}(t_{n+1})\,\Phi(t_{n+1}, t_n) \qquad (34.26)$$

et la condition à l'instant t_N

$$\mathcal{S}(t_N) = \mathcal{F}. \qquad (34.27)$$

La relation (8)

$$\mathcal{S}(t_n) = [\Phi(t_{n+1}, t_n) - L(t_n)\,\mathcal{B}(t_n)]^*\,\mathcal{S}(t_{n+1}) \times \\ \times [\Phi(t_{n+1}, t_n) - L(t_n)\,\mathcal{B}(t_n)] + \mathcal{B}^*(t_n)\,\mathfrak{N}_n\mathcal{B}(t_n) + \mathfrak{M}_n$$

implique que si la matrice $\mathcal{S}(t_{n+1})$ est définie non négative, $\mathcal{S}(t_n)$ est également une matrice définie non négative.

Du fait que $\mathcal{F}$ est une matrice définie non négative, conformément à (27) $\mathcal{S}(t_N)$ est également une matrice définie non négative. De cette façon la matrice symétrique $\mathcal{S}(t_n)$ du type $n \times n$, $0 \leqslant n \leqslant N$ est une matrice définie non négative.

Montrons maintenant que

$$x^*(t_N)\,\mathcal{F}x(t_N) + \sum_{n=0}^{N-1} [x^*(t_n)\,\mathfrak{M}_n x(t_n) + u_n^*\mathfrak{N}_n u_n] = \\ = x^*(t_0)\,\mathcal{S}(t_0)\,x(t_0) + \sum_{n=0}^{N-1} [u_n + \mathcal{B}(t_n)\,x(t_n)]^*\,[L^*(t_n) \times \\ \times \mathcal{S}(t_{n+1})\,L(t_n) + \mathfrak{N}_n]\,[u_n + \mathcal{B}(t_n)\,x(t_n)] + \sum_{n=0}^{N-1} \{w_n^*\Gamma^*(t_n) \times \\ \times \mathcal{S}(t_{n+1})\,[\Phi(t_{n+1}, t_n)\,x(t_n) + L(t_n)\,u_n] + \\ + [\Phi(t_{n+1}, t_n)\,x(t_n) + L(t_n)\,u_n]^*\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,w_n + \\ + w_n^*\Gamma^*(t_n)\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,w_n\}. \qquad (34.28)$$

La relation (28) peut être justifiée de la façon suivante. En vertu de (27) on a l'identité

$$x^*(t_N)\,\mathcal{F}x(t_N) = x^*(t_N)\,\mathcal{S}(t_N)\,x(t_N) = \\ = x^*(t_0)\,\mathcal{S}(t_0)\,x(t_0) + \sum_{n=0}^{N-1} x^*(t_{n+1})\mathcal{S}(t_{n+1})\,x(t_{n+1}) - \\ - \sum_{n=0}^{N-1} x^*(t_n)\,\mathcal{S}(t_n)\,x(t_n),$$

ou

$$x^*(t_N)\,\mathcal{F}x(t_N) = x^*(t_0)\,\mathcal{S}(t_0)\,x(t_0) + \\ + \sum_{n=0}^{N-1} [x^*(t_{n+1})\,\mathcal{S}(t_{n+1})\,x(t_{n+1}) - x^*(t_n)\,\mathcal{S}(t_n)\,x(t_n)]. \qquad (34.29)$$

D'après l'équation (1),

$$x^* (t_{n+1}) \mathscr{S} (t_{n+1}) x (t_{n+1}) =$$
$$= [\Phi (t_{n+1}, t_n) x (t_n) + L (t_n) u_n + \Gamma (t_n) w_n]^* \mathscr{S} (t_{n+1}) \times$$
$$\times [\Phi (t_{n+1}, t_n) x (t_n) + L (t_n) u_n + \Gamma (t_n) w_n]. \quad (34.30)$$

D'après l'équation (8)

$$x^* (t_n) \mathscr{S} (t_n) x (t_n) =$$
$$= x^* (t_n) \{\Phi^* (t_{n+1}, t_n) \mathscr{S} (t_{n+1}) \Phi (t_{n+1}, t_n) + \mathfrak{M}_n -$$
$$- \mathscr{B}^* (t_n) [\mathfrak{N}_n + L^* (t_n) \mathscr{S} (t_{n+1}) L (t_n)] \mathscr{B} (t_n)\} x (t_n). \quad (34.31)$$

En portant les expressions (30) et (31) dans (29), on amène sans peine l'expression (29) à la forme (28)

Notons que dans le cas où $\Gamma (t_n) \equiv 0$, l'expression (28) devient

$$x^* (t_N) \mathscr{F} x (t_N) + \sum_{n=0}^{N-1} [x^* (t_n) \mathfrak{M}_n x (t_n) + u_n^* \mathfrak{N}_n u_n] =$$
$$= x^* (t_0) \mathscr{S} (t_0) x (t_0) + \sum_{n=0}^{N-1} [u_n + \mathscr{B} (t_n) x (t_n)]^* \times$$
$$\times [L^* (t_n) \mathscr{S} (t_{n+1}) L (t_n) + \mathfrak{N}_n] [u_n + \mathscr{B} (t_n) x (t_n)] \quad (34.32)$$

Comme nous l'avons montré plus haut, la matrice symétrique $\mathscr{S} (t_n)$ du type $n \times n$, $0 \leqslant n \leqslant N$, est définie non négative. C'est pourquoi le deuxième terme du second membre de l'expression (32) est non négatif. Comme il découle de l'équation aux différences (26), la matrice $\mathscr{S} (t_n)$, $0 \leqslant n \leqslant N$ ne dépend pas de la commande u_n Ainsi, la relation (32) entraîne l'inégalité

$$x^* (t_N) \mathscr{F} x (t_N) + \sum_{n=0}^{N-1} [x^* (t_n) \mathfrak{M}_n x (t_n) + u_n^* \mathfrak{N}_n u_n] \geqslant$$
$$\geqslant x^* (t_0) \mathscr{S} (t_0) x (t_0), \quad (34.33)$$

le signe d'égalité de (33) n'ayant lieu que pour la commande u_n $(n = 0, 1, \ldots, N - 1)$, déterminée par la formule (6):

$$u_n = u_n^0 = - \mathscr{B}(t_n) x (t_n).$$

Ici la matrice $\mathscr{B} (t_n)$ de type $p \times n$ est définie par l'expression (7); de plus, on suppose que la matrice $\mathscr{N}_n$ qui fait partie de l'expression (7)

$$\mathscr{N}_n = \mathfrak{N}_n + L^* (t_n) \mathscr{S} (t_{n+1}) L (t_n), \qquad n = 0, 1 \ldots N - 1 \quad (34.34)$$

est définie positive. Le choix de la matrice $\mathfrak{N}_n$ définie positive de type $p \times p$ dans la fonctionnelle (5) assure l'observation de cette contrainte.

Comme il s'ensuit de (33) et (22), dans le cas où $\Gamma(t_n) \equiv 0$, la valeur minimale de la fonctionnelle (5) est

$$\min J = \mathrm{M}\,[x^*(t_0)\,\mathcal{S}(t_0)\,x(t_0)] = \\ = \mu^*\mathcal{S}(t_0)\,\mu + \mathrm{Sp}\,[\mathcal{S}(t_0)\,P(t_0)]. \qquad (34.35)$$

Dans le cas général, où $\Gamma(t_n) \not\equiv 0$, d'après (28), la fonctionnelle (5) devient

$$J = \mathrm{M}\left\{x^*(t_N)\,\mathcal{F}x(t_N) + \sum_{n=0}^{N-1}[x^*(t_n)\,\mathfrak{M}_n x(t_n) + u_n^*\mathfrak{N}_n u_n]\right\} = \\ = \mathrm{M}\left\{x^*(t_0)\,\mathcal{S}(t_0)\,x(t_0) + \sum_{n=0}^{N-1} w_n^*\Gamma^*(t_n)\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,w_n + \right. \\ + \sum_{n=0}^{N-1}[u_n + \mathcal{B}(t_n)\,x(t_n)]^*\,[L^*(t_n)\,\mathcal{S}(t_{n+1}) \times \\ \left. \times L(t_n) + \mathfrak{N}_n]\,[u_n + \mathcal{B}(t_n)\,x(t_n)]\right\}. \qquad (34.36)$$

Conformément à (22) et compte tenu de $\mathrm{M}\,[w_n] = 0$, on a

$$\mathrm{M}\,[x^*(t_0)\,\mathcal{S}(t_0)\,x(t_0)] = \mu^*\mathcal{S}(t_0)\,\mu + \mathrm{Sp}\,[\mathcal{S}(t_0)\,P(t_0)], \\ \mathrm{M}\,[w_n^*\Gamma^*(t_n)\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,w_n] = \mathrm{Sp}\,[\Gamma^*(t_n)\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,Q_n]. \qquad (34.37)$$

Ainsi, l'expression (36) devient

$$J = \mathrm{M}\left\{x^*(t_N)\,\mathcal{F}x(t_N) + \sum_{n=0}^{N-1}[x^*(t_n)\,\mathfrak{M}_n x(t_n) + u_n^*\mathfrak{N}_n u_n]\right\} = \\ = \mu^*\mathcal{S}(t_0)\,\mu + \mathrm{Sp}\,[\mathcal{S}(t_0)\,P(t_0)] + \sum_{n=0}^{N-1}\mathrm{Sp}\,[\Gamma^*(t_n)\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,Q_n] + \\ + \mathrm{M}\left\{\sum_{n=0}^{N-1}[u_n + \mathcal{B}(t_n)\,x(t_n)]^*\,[L^*(t_n)\,\mathcal{S}(t_{n+1})\,L(t_n) + \right. \\ \left. + \mathfrak{N}_n]\,[u_n + \mathcal{B}(t_n)\,x(t_n)]\right\}. \qquad (34.38)$$

D'une façon analogue à (33), la relation (38) entraînc l'inégalité suivante:

$$J = \mathrm{M}\left\{x^*(t_N)\,\mathcal{F}x(t_N) + \sum_{n=0}^{N-1}[x^*(t_n)\,\mathfrak{M}_n x(t_n) + u_n^*\mathfrak{N}_n u_n]\right. \geqslant \\ \geqslant \mu^*\mathcal{S}(t_0)\,\mu + \mathrm{Sp}\,[\mathcal{S}(t_0)\,P(t_0)] + \\ + \sum_{n=0}^{N-1}\mathrm{Sp}\,[\Gamma^*(t_n)\,\mathcal{S}(t_{n+1})\,\Gamma(t_n)\,Q_n] \qquad (34.39)$$

En outre, dans (39) le signe d'égalité n'a lieu que pour la commande u_n ($n = 0, 1, \ldots, N-1$) définie par la formule (6)

$$u_n = u_n^0 = -\mathscr{B}(t_n)\, x(t_n),$$

où la matrice $\mathscr{B}(t_n)$ d'ordre $p \times n$ est définie par l'expression (7).

Ici, de même que précédemment (34), on observe la condition

$$\mathscr{N}_n = \mathfrak{N}_n + L^*(t_n)\,\mathscr{S}(t_{n+1})\,L(t_n) > 0, \quad n = 0, 1, \ldots, N-1, \tag{34.40}$$

c'est-à-dire la condition d'une matrice $\mathscr{N}_n$ définie positive.

D'après (39), la valeur minimale de la fonctionnelle (5) est déterminée par l'expression (10). Le théorème est démontré.

2. Systèmes linéaires à critère quadratique de performance avec information incomplète sur l'état du système. Théorème de séparation.

Considérons le système décrit par l'équation vectorielle aux différences finies

$$x(t_{n+1}) = \Phi(t_{n+1}, t_n)\, x(t_n) + L(t_n)\, u_n + \Gamma(t_n)\, w_n, \qquad n = 0, 1, 2, \ldots, \quad x(t_0) = s. \tag{34.41}$$

Ici x est le vecteur d'état de dimension n; $\Phi(t_{n+1}, t_n)$, la matrice $n \times n$; $L(t_n)$, la matrice $n \times p$; u_n, le vecteur de commande de dimension p; $\Gamma(t_n)$, la matrice $n \times r$; w_n, la suite vectorielle de dimension r des variables aléatoires gaussiennes réciproquement indépendantes à moyenne nulle.

Les mesures permettent de déterminer le vecteur z_n de dimension m, $n = 0, 1, 2, \ldots$, défini par la relation

$$z_n = H_n x(t_n) + v_n, \qquad n = 0, 1, 2, \ldots, \tag{34.42}$$

où H_n est la matrice $m \times n$; v_n, la suite vectorielle de dimension m des variables aléatoires gaussiennes réciproquement indépendantes à moyenne nulle. La mesure directe de $x(t_n)$ est supposée impossible.

Les matrices de corrélation des suites discontinues w_n et v_n sont de la forme

$$\begin{aligned} \mathrm{M}[w_n w_k^*] &= Q_n \delta_{nk}, \\ \mathrm{M}[v_n v_k^*] &= R_n \delta_{nk}, \\ \mathrm{M}[w_n v_k^*] &= 0, \end{aligned} \tag{34.43}$$

où Q_n est la matrice symétrique définie non négative $r \times r$, et R_n, la matrice symétrique définie positive $m \times m$; par ailleurs, $R_n \geqslant \alpha E$, où $\alpha > 0$ et E est la matrice unité.

δ_{nk} désigne le symbole de Kronecker

$$\delta_{nk} = \begin{cases} 1 & \text{pour} \quad n = k, \\ 0 & \text{pour} \quad n \neq k. \end{cases}$$

Par s on désigne le vecteur aléatoire gaussien de dimension n à valeur moyenne

$$M[s] = \mu \tag{34.44}$$

et à matrice de corrélation

$$M[(s-\mu)(s-\mu)^*] = P(t_0). \tag{34.45}$$

Ici $P(t_0)$ est la matrice définie non négative de type $n \times n$ qu'on suppose connue. De plus, on suppose que $w(t)$, $v(t)$ et s sont non corrélés et, par conséquent, indépendants.

Par $\hat{x}(t_n \mid j)$ on désigne l'espérance mathématique conditionnelle de $x(t_n)$ avec $z_0, z_1, \ldots, z_j$ donnés:

$$\begin{gathered}\hat{x}(t_n \mid j) = M[x(t_n) \mid z_0, z_1, \ldots, z_j], \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \hat{x}(t_n \mid n) = M[x(t_n) \mid z_0, z_1, \ldots, z_j, \ldots, z_n].\end{gathered} \tag{34.46}$$

De plus, introduisons les notations

$$\tilde{x}(t_n \mid j) = x(t_n) - \hat{x}(t_n \mid j), \tag{34.47}$$

$$P(t_n \mid n) = M[\tilde{x}(t_n \mid n)\, \tilde{x}^*(t_n \mid n)], \tag{34.48}$$

$$P(t_n \mid n-1) = M[\tilde{x}(t_n \mid n-1)\, \tilde{x}^*(t_n \mid n-1)]. \tag{34.49}$$

De même qu'au point 1, considérons le problème de régulation de l'état du système, c'est-à-dire celui dans lequel le but de la commande est de maintenir les moyennes des coordonnées de phase $x_j(t_n)$ dans le voisinage du zéro. Considérons que l'intervalle de temps (t_0, t_N) dans lequel on réalise la commande est fixé. Ainsi, de même qu'au point 1, la commande u_n, $n = 0, 1, \ldots, N-1$ (où N est fixé) doit être choisie de façon à minimiser la fonctionnelle

$$J = M\left\{\langle x(t_N), \mathcal{F}x(t_N)\rangle + \sum_{n=0}^{N-1} [\langle x(t_n), \mathfrak{M}_n x(t_n)\rangle + \langle u_n, \mathfrak{N}_n u_n\rangle]\right\}. \tag{34.50}$$

Ici $\mathcal{F}$ et $\mathfrak{M}_n$ sont des matrices symétriques définies non négatives de type $n \times n$, et $\mathfrak{N}_n$, la matrice symétrique définie positive de type $p \times p$.

En posant que les commandes admissibles u_n sont des fonctions de $z_0, z_1, \ldots, z_{n-1}$ et t_n, désignons par u_n^0 la commande optimale, c'est-à-dire la commande minimisant la fonctionnelle (50).

Théorème 2. *Dans le problème à extrémité libre de la trajectoire et à temps fixé* (41), (42), (50), *avec une information incomplète sur l'état du système, la commande optimale* u_n^0 *s'écrit*

$$u_n^0 = -\mathcal{B}(t_n)\, \hat{x}(t_n \mid n-1). \tag{34.51}$$

Ici $\hat{x}(t_n \mid n-1)$ *est déterminée par les relations récurrentes*

$$\hat{x}(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n)\,\hat{x}(t_n \mid n) + L(t_n)\,u_n^0, \tag{34.52}$$

$$\hat{x}(t_n \mid n) = \hat{x}(t_n \mid n-1) + A_n\,[z_n - H_n\,\hat{x}(t_n \mid n-1)],$$
$$\hat{x}(t_0 \mid -1) = \hat{x}(t_0) = \mu \tag{34.53}$$

où

$$A_n = P(t_n \mid n-1)\,H_n^*\,[H_n P(t_n \mid n-1)\,H_n^* + R_n]^{-1}, \tag{34.54}$$

$$P(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n)\,P(t_n \mid n)\,\Phi^*(t_{n+1}, t_n) + \\ + \Gamma(t_n)\,Q_n\Gamma^*(t_n), \tag{34.55}$$

$$P(t_n \mid n) = (E - A_nH_n)P(t_n \mid n-1)(E - A_nH_n)^* + \\ + A_nR_nA_n^* = P(t_n \mid n-1) - A_nH_nP(t_n \mid n-1), \tag{34.56}$$
$$P(t_0 \mid -1) = P(t_0) = \mathbf{M}\,[(s-\mu)(s-\mu)^*].$$

Les relations récurrentes (52) *et* (53) *sont équivalentes à l'équation vectorielle aux différences*

$$\hat{x}(t_{n+1} \mid n) = \Phi(t_{n+1}, t_n)\,\hat{x}(t_n \mid n-1) + \\ + L(t_n)\,u_n^0 + \Phi(t_{n+1}, t_n)\,P(t_n \mid n-1) \times \\ \times H_n^*\,[H_nP(t_n \mid n-1)\,H_n^* + R_n]^{-1}\,[z_n - H_n\hat{x}(t_n \mid n-1)], \tag{34.57}$$
$$\hat{x}(t_0 \mid -1) = \hat{x}(t_0) = \mu.$$

La matrice $\mathscr{B}(t_n)$ *de type* $p \times n$ *qui fait partie de la formule* (51) *est de la forme*

$$\mathscr{B}(t_n) = [\mathfrak{N}_n + L^*(t_n)\,\mathscr{S}(t_{n+1})\,L(t_n)]^{-1} \times \\ \times L^*(t_n)\,\mathscr{S}(t_{n+1})\,\Phi(t_{n+1}, t_n) \tag{34.58}$$

où la matrice $\mathscr{S}(t_n)$ *de type* $n \times n$ *est définie par la relation de récurrence*

$$\mathscr{S}(t_n) = \Phi^*(t_{n+1}, t_n)\,\mathscr{S}(t_{n+1})\,\Phi(t_{n+1}, t_n) + \\ + \mathfrak{M}_n - \mathscr{B}^*(t_n)\,[\mathfrak{N}_n + L^*(t_n)\,\mathscr{S}(t_{n+1})\,L(t_n)]\,\mathscr{B}(t_n) = \\ = [\Phi(t_{n+1}, t_n) - L(t_n)\,\mathscr{B}(t_n)]^*\,\mathscr{S}(t_{n+1})\,\Phi(t_{n+1}, t_n) + \mathfrak{M}_n = \\ = [\Phi(t_{n+1}, t_n) - L(t_n)\,\mathscr{B}(t_n)]^*\,\mathscr{S}(t_{n+1})\,[\Phi(t_{n+1}, t_n) - L(t_n)\,\mathscr{B}(t_n)] + \\ + \mathscr{B}^*(t_n)\,\mathfrak{N}_n\mathscr{B}(t_n) + \mathfrak{M}_n \tag{34.59}$$

et par la condition à l'instant t_N

$$\mathscr{S}(t_N) = \mathscr{F}. \tag{34.60}$$

La valeur minimale de la fonctionnelle (50) *s'écrit*

$$\min J = \mu^* \mathscr{S}(t_0)\,\mu + \mathrm{Sp}\,[\mathscr{S}(t_0)\,P(t_0)] + \sum_{n=0}^{N-1} \mathrm{Sp}\,[\Gamma^*(t_n)\,\mathscr{S}(t_{n+1})\,\Gamma(t_n)\,Q_n] + \sum_{n=0}^{N-1} \mathrm{Sp}\,[P(t_n\,|\,n-1)\,\mathscr{B}^*(t_n)\,L^*(t_n)\,\mathscr{S}(t_{n+1})\,\Phi(t_{n+1}, t_n)]. \quad (34.61)$$

D é m o n s t r a t i o n d u t h é o r è m e 2. Comme nous l'avons dit plus haut, la mesure directe de l'état du système étant impossible, on adopte comme commandes admissibles u_n les fonctions de $z_0, z_1, \ldots, z_{n-1}, t_n$. Puisque les équations (1), (7), (8) et (9) qui ont lieu dans le problème à information complète sur l'état du système coïncident respectivement avec les équations (41), (58), (59) et (60), les expressions (28) et (38) sont également valables pour le problème à information incomplète sur l'état du système.

Toutefois, puisque la commande définie par la formule (6) n'appartient pas au nombre de commandes admissibles du problème avec information incomplète, le dernier terme de l'expression (38)

$$J_1 = \mathrm{M}\left\{ \sum_{n=0}^{N-1} [u_n + \mathscr{B}(t_n)\,x(t_n)]^*\,[L^*(t_n)\,\mathscr{S}(t_{n+1})\,L(t_n) + \mathfrak{N}_n] \times [u_n + \mathscr{B}(t_n)\,x(t_n)] \right\} \quad (34.62)$$

est maintenant différent du zéro.

En appliquant le lemme 2 à la fonction (62), on obtient conformément à (16)

$$\min_{u_n} \mathrm{M}\left\{ \sum_{n=0}^{N-1} [u_n + \mathscr{B}(t_n)\,x(t_n)]^*\,[L^*(t_n)\,\mathscr{S}(t_{n+1})\,L(t_n) + \mathfrak{N}_n] \times [u_n + \mathscr{B}(t_n)\,x(t_n)] \right\} = \mathrm{M}_{\mathfrak{Z}}\left(\min_{u_n} \mathrm{M}\left\{ \sum_{n=0}^{N-1} [u_n + \mathscr{B}(t_n)\,x(t_n)]^* \times [L^*(t_n)\,\mathscr{S}(t_{n+1})\,L(t_n) + \mathfrak{N}_n]\,[u_n + \mathscr{B}(t_n)\,x(t_n)] \,\Big|\, \mathfrak{Z}_{n-1} \right\} \right) \quad (34.63)$$

où $\mathfrak{Z}_{n-1}$ est le vecteur dont les éléments sont des vecteurs $z_0, z_1, \ldots, z_{n-1}$.

Ici $\mathrm{M}\{\cdot\,|\,\mathfrak{Z}_{n-1}\}$ désigne l'espérance mathématique conditionnelle pour $\mathfrak{Z}_{n-1}$ donné, c'est-à-dire pour $z_0, z_1, \ldots, z_{n-1}$ donnés. La minimisation par rapport à u_n se fait sur l'ensemble des commandes u_n qui sont des fonctions de $z_0, z_1, \ldots, z_{n-1}$. Le symbole $\mathrm{M}_{\mathfrak{Z}}$ désigne l'espérance matématique d'après la distribution commune des variables aléatoires $z_0, z_1, \ldots, z_{n-1}$.

Compte tenu des notations introduites dans (46) et (49),

$$\mathrm{M}\,[x(t_n)\,|\,z_0, z_1, \ldots, z_{n-1}] = \hat{x}(t_n\,|\,n-1),$$

$$\mathrm{cov}\,[x(t_n)\,|\,z_0, z_1, \ldots, z_{n-1}] = P(t_n\,|\,n-1),$$

on a en vertu du lemme 3 la relation

$$\mathbf{M}\left\{\sum_{n=0}^{N-1}[u_n+\mathscr{B}(t_n)x(t_n)]^*[L^*(t_n)\mathscr{S}(t_{n+1})L(t_n)+\right.$$
$$\left.+\mathfrak{N}_n][u_n+\mathscr{B}(t_n)x(t_n)]\mid\mathfrak{Z}_{n-1}\right\}=$$
$$=\sum_{n=0}^{N-1}\mathrm{Sp}\{\mathscr{B}^*(t_n)[L^*(t_n)\mathscr{S}(t_{n+1})L(t_n)+\mathfrak{N}_n]\mathscr{B}(t_n)P(t_n\mid n-1)\}+$$
$$+\sum_{n=0}^{N-1}[u_n+\mathscr{B}(t_n)\hat{x}(t_n\mid n-1)]^*[L^*(t_n)\mathscr{S}(t_{n+1})L(t_n)+$$
$$+\mathfrak{N}_n][u_n+\mathscr{B}(t_n)\hat{x}(t_n\mid n-1)]. \qquad (34.64)$$

Les relations récurrentes (55) et (56) montrent que la matrice de corrélation $P(t_n\mid n-1)$ est indépendante de u. Ainsi, comme il résulte de (64) et (63), la commande optimale u_n^0 est déterminée par la formule (51):

$$u_n^0=-\mathscr{B}(t_n)\hat{x}(t_n\mid n-1). \qquad (34.65)$$

Comme il s'ensuit de (38), (63), (64) et (65), la valeur minimale de la fonctionnelle (50) est

$$\min J=\mu^*\mathscr{S}(t_0)\mu+\mathrm{Sp}[\mathscr{S}(t_0)P(t_0)]+$$
$$+\sum_{n=0}^{N-1}\mathrm{Sp}[\Gamma^*(t_n)\mathscr{S}(t_{n+1})\Gamma(t_n)Q_n]+$$
$$+\sum_{n=0}^{N-1}\mathrm{Sp}[P(t_n\mid n-1)\mathscr{B}^*(t_n)[L^*(t_n)\mathscr{S}(t_{n+1})L(t_n)+\mathfrak{N}_n]\mathscr{B}(t_n).$$
$$(34.66)$$

En remplaçant la matrice $\mathscr{B}(t_n)$ par l'expression (58), on ramène la formule (66) à la forme (61).

Les relations (52) à (56) données dans le théorème pour $u_n=0$ sont déduites au § 31. Au § 33 (théorème 1), on a étudié le cas où $u_n\neq 0$. Le théorème est démontré

BIBLIOGRAPHIE *

1. Isaacs, R., *Differential Games.* New York, Wiley, 1965.
2. Айзерман М. А., *Теория автоматического управления.* Изд-во « Наука », 1966.
3. Айзерман М. А., Гантмахер Ф. Р., *Абсолютная устойчивость регулируемых систем.* Изд-во Академии Наук СССР, 1963.
4. Андронов А. А., Витт А. А., Хайкин С. Э., *Теория колебаний.* Изд. 2-е, Физматгиз, 1959.
5. Athans, M., A direct derivation of the Optimal Linear Filters, using the maximal principle. IEEE, Transactions on Automatic Control, № 6, 1967, pp. 690-697.
6. Athans, M., Falb P. L., *Optimal Control.* Mc Graw Hill, 1966.
7. Барбашин Е. А., *Введение в теорию устойчивости.* Изд-во « Наука », 1967.
8. Bellman, R., *Dynamic Programming.* Princeton University Press, N.Y. 1957.
9. Bellman, R., Glicksberg, I., Gross, O., *Some aspects of the mathematical theory of control processus.* Santa Monica, Rand corp., 1958.
10. Bellman R., Kalaba R., *Dynamic programming and modern control theory.* New York-London, 1965.
11. Bellman, R., Cooke, K., Differential-difference equations. New York-London, Acad. Press, 1963.
12. Боголюбов Н. Н., Митропольский Ю. А., *Асимптотические методы в теории нелинейных колебаний.* Гостехиздат, 1955.
13. Bode, H., *Network Analysis and feedback amplifier design.* New York, 1946.
14. Болтянский В. Г., *Достаточные условия оптимальности и обоснование метода динамического программирования.* Изв. АН СССР, серия математическая, т. 28, № 3, 1964, стр. 481-514.
15. Болтянский В. Г., *Математические методы оптимального управления.* Изд. 2-е, изд-во « Наука », 1969.
16. Bryson, A., Yu-Chi Ho, *Applied Optimal Control.* Blaisdell Publishing Company. 1969.
17. Булгаков Б. В., *Колебания.* Гостехиздат, 1954.
18. Бутковский А. Г., *Теория оптимального управления системами с распределенными параметрами.* Изд-во « Наука », 1965.
19. Bucy R. S., Joseph P. D., *Filtering for Stochastic Processes with Applications to Guidance.* Interscience Publishers, New York, 1968.
20. Wiener N., *Extrapolation, interpolation and smoothing of stationary Time Series.* J. Wiley, New York, Second printing, 1950.
21. Гантмахер Ф. Р., *Теория матриц.* Изд-во « Наука », 1966.

* Cette liste bibliographique n'a aucune prétention d'être complète. Nous n'avons indiqué que les monographies principales, les tours d'horizon et un petit nombre d'articles.

22. Гантмахер Ф. Р., Якубович В. А., *Абсолютная устойчивость нелинейных регулируемых систем*. Труды II Всесоюзного съезда по теоретической и прикладной механике. Обзорные доклады, вып. 1, Изд-во « Наука », 1965, стр. 30-63.
23. Гельфанд И. М., *Лекции по линейной алгебре*. Изд-во « Наука », 1966.
24. Гельфанд И. М., Фомин С. В., *Вариационное исчисление*. Физматгиз, 1965.
25. Гихман И. И., Скороход А. В., *Введение в теорию случайных процессов*. Изд-во « Наука », 1965.
26. Гихман И. И., Скороход А. В., *Стохастические дифференциальные уравнения*. Изд-во « Наукова думка », 1967.
27. Гноенский А. С., Каменский А. Г., Эльсгольц Л. Э., *Математические основы теории управляемых систем*. Изд-во « Наука », 1969.
28. Демидович Б. П., *Лекции по математической теории устойчивости*. Изд-во « Наука », 1967.
29. Демьянов В. Ф., Рубинов А. М., *Приближенные методы решения экстремальных задач*. Изд-во Ленинградского университета, 1968.
30. Дубовицкий А. Я., Милютин А. А., *Задачи на экстремум при наличии ограничений*. Журнал вычислительной математики и математической физики, т. 5, № 3, 1965, стр. 395-493.
31. Зубов В. И., *Колебания в нелинейных управляемых системах*. Судпромгиз, 1962.
32. Ишлинский А. Ю., *Инерциальное управление баллистическими ракетами*. Изд-во « Наука », 1968.
33. Kalman, R., *A new approach to linear filtering and prediction problems*. J. Basic Engr. (ASME Trans.), v. 82, D, 1960, pp. 35-45.
34. Kalman R., *Contributions to the Theory of optimal control*. Boletin de la Sociedad Matematica Mexicana, v. 5, Segunda serie № 1, 1960, pp. 102-119.
35. Kalman, R., *New methods in Wiener filtering theory*. Proceedings of the First Symposium on Engineering applications of random function theory and Probability, J. Wiley, 1963, pp. 270-388.
36. Kalman, R., *On the general theory of control systems*. Proc. I.F.A.C., Moscow, 1960, Butterworts, 1961.
37. Kalman, R., Bucy R., *New results in linear filtering and prediction theory*. J. Basic Engr. (ASME Trans.), v. 83, D, 1961, pp. 95-108.
38. Канторович Л. В., Акилов Г. П., *Функциональный анализ в нормированных пространствах*, 1959, стр. 175.
39. Кирин Н. Е., *Вычислительные методы в теории оптимального управления*. Изд-во Ленинградского университета, 1968.
40. Колмогоров А. Н., *Интерполяция и экстраполяция стационарных случайных последовательностей*. Изв. АН СССР, серия математическая, № 5, 1941.
41. Конторович М. И., *Операционное исчисление и процессы в электрических цепях*. Изд. 3-е, изд-во « Наука », 1964.
42. Красовский Н. Н., *Игровые задачи о встрече движений*. Изд-во « Наука », 1970.
43. Красовский Н. Н., *Некоторые задачи теории устойчивости движения*. Физматгиз, 1959.
44. Красовский Н. Н., *Теория оптимальных управляемых систем*. *Механика в СССР за 50 лет*, т. I. Изд-во « Наука », 1968, стр. 178-244.
45. Красовский Н. Н., *Теория управления движением*. Изд-во « Наука », 1968.
46. Кротов В. Ф., *Методы решения вариационных задач на основе достаточных условий абсолютного минимума*, ч. I, II, III. Автоматика и телемеханика, т. 23, № 12, 1962; т. 24, № 5, 1963; т. 25, № 7, 1964.
47. Крылов Н. М., Боголюбов Н. Н., *Введение в нелинейную механику*. Изд-во АН УССР, 1937.
48. Kulikowski, R., *Procesy aptymalne i adaptacyjne w ukadach regulacji automatycznej*. Warszawa, Wroclaw, 1965.

49. Kushner H. J., *Stochastic stability and control.* Academic Press, New York-London, 1967.
50. Лаппо-Данилевский И. А., *Применение функций от матриц к теории линейных систем обыкновенных дифференциальных уравнений.* Гостехиздат, 1957.
51. Lattes, R., Lions, J.-L., *Méthode de quasi-réversibilité et applications.* Dunod, Paris, 1967.
52. Летов А. М., *Аналитическое конструирование регуляторов.* « Автоматика и телемеханика », т. 21, № 4, № 5, № 6, 1960; т. 22, № 4, 1961; т. 23, № 11, 1962.
53. Летов А. М., *Динамика полета и управление.* Изд-во « Наука », 1969.
54. Lefschetz, S., *Stability of nonlinear control systems.* Academic Press, New York-London, 1965.
55. Lefschetz, S., La Salle J., *Stability by Liapounov's direct method,* with applications. New York-London, Academic Press, 1961.
56. Lee, E. B., Markus, L., *Foundations of optimal control theory,* J. Wiley, 1967.
57. Lee, R., *Optimal estimation, identification and control.* Cambridge (Mass.), MIT press, 1964.
58. Lions, J.-L., *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles.* Dunod, Gauthier-Villars, 1968.
59. Лурье А. И., *Некоторые нелинейные задачи теории автоматического регулирования.* Гостехиздат, 1951.
60. Liapounov, A., *Le problème général de la stabilité du mouvement.* Faculté des Sciences de Toulouse Annales, vol. 9, 1907).
61. Mc Kinsey, J., *Introduction to the theory of games.* Mc Graw Hill, New York, 1952.
62. Малкин И. Г., *Теория устойчивости движения.* Изд. 2-е, изд-во « Наука », 1966.
63. Leitmann, G. (ed.), *Optimization techniques, with applications to aerospace systems.* New York-London, Acad. press, 1962.
64. Митропольский Ю. А., *Проблемы асимптотической теории нестационарных колебаний.* Изд-во « Наука », 1964.
65. Михайлов А. В., *Метод гармонического анализа в теории регулирования.* « Автоматика и телемеханика », № 3, 1938, стр. 27-81.
66. Моисеев Н. Н., *Численные методы оптимального управления.* Вычислительный центр АН СССР, 1968.
67. Мышкис А. Д., *Линейные дифференциальные уравнения с запаздывающим аргументом.* Гостехиздат, 1951.
68. Немыцкий В. В., Степанов В. В., *Качественная теория дифференциальных уравнений.* Гостехиздат, 1949.
69. *Основы автоматического управления.* Под редакцией В. С. Пугачева. Физматгиз, 1963.
70. Охоцимский Д. Е., Энеев Т. М., *Некоторые вариационные задачи, связанные с запуском искусственного спутника Земли.* Успехи физических наук, т. 63, № 1а, 1957.
71. Понтрягин Л. С., *К теории дифференциальных игр.* Успехи математических наук, т. XXI, вып. 4 (130), 1966, стр. 219-275.
72. Понтрягин Л. С., Болтянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф., *Математическая теория оптимальных процессов.* Физматгиз, 1961, изд. 2-е, изд-во « Наука », 1969.
73. Пóпов В.-М., *Об абсолютной устойчивости нелинейных систем автоматического регулирования.* « Автоматика и телемеханика », т. 22, № 8, 1961.
74. Пшеничный Б. Н., *Необходимые условия экстремума,* Изд-во « Наука », 1969.
75. Poincaré, H., *Sur les courbes définies par des équations différentielles.* Journal des Mathématiques pures et appliquées, 1881-1882; réédité en 1951 par Gauthier-Villars: *Œuvres,* tome I.

76. Пугачев В. С., *Теория случайных функций и ее приложение к задачам автоматического управления.* Физматгиз, 1960.
77. Ройтенберг Я. Н., *Некоторые задачи управления движением.* Физматгиз, 1963.
78. Ройтенберг Я. Н., *Теория нелинейных колебаний и автоматическое регулирование*, ч. I, II. Изд-во Московского Университета, 1967.
79. Румянцев В. В., *Метод функций Ляпунова в теории устойчивости движения. Механика в СССР за 50 лет.* т. I. « Общая и прикладная механика ». Изд-во « Наука », 1968, стр. 7-66.
80. Свешников А. А., *Прикладные методы теории случайных функций.* Изд-во « Наука », 1968.
81. Стратонович Р. Л., *Условные марковские процессы и их применение в теории оптимального управления.* Изд-во МГУ, 1966.
82. Titchmarsh, E. C., *Introduction to the theory of Fourier integrals.* Oxford, The Clarendon Press, 1948.
83. Тихонов А. Н., *О методах регуляризации задач оптимального управления.* ДАН СССР, т. 162, №,4, 1965.
84. Троицкий В. А., *О вариационных методах оптимизации процессов управления.* Прикладная математика и механика, т. 26, № 1, 1962.
85. Feldbaum, A., *Principes théoriques des systèmes asservis optimaux.* Editions MIR, 1973.
86. Филиппов А. Ф., *О некоторых вопросах теории оптимального регулирования.* Вестник МГУ, серия матем., мех., астр., физ., хим., № 2, 1959.
87. Хасьминский Р. З., *Устойчивость систем дифференциальных уравнений при случайных возмущениях их параметров.* Изд-во « Наука », 1969.
88. Цыпкин Я. З., *Адаптация и обучение в автоматических системах.* Изд-во « Наука », 1968.
89. Четаев Н. Г., *Устойчивость движения.* Изд. 2-е, Гостехиздат, 1955.
90. Шилов Г. Е., *Математический анализ.* Специальный курс. Физматгиз, 1961.
91. Шилов Г. Е., *Математический анализ.* Второй специальный курс. Изд-во « Наука », 1965.
92. Эльсгольц Л. Э., *Введение в теорию дифференциальных уравнений с отклоняющимся аргументом.* Изд-во « Наука », 1964.
93. Якубович В. А., *К теории адаптивных систем.* ДАН СССР, т. 182, № 3, 1968, стр. 518-521.
94. Якубович В. А., *Метод матричных неравенств в теории устойчивости нелинейных регулируемых систем*, ч. I, II. Автоматика и телемеханика, т. 25, № 7, 1964; т. 26, № 4, 1965.
95. Якубович В. А., *Решение некоторых специальных матричных неравенств, встречающихся в нелинейной теории регулирования.* ДАН СССР, т. 156, № 2, 1964, стр. 278-281.

INDEX ALPHABÉTIQUE

TABLE DES MATIÈRES

À NOS LECTEURS

Les Editions Mir vous seraient très reconnaissantes de bien vouloir leur communiquer votre opinion sur le contenu de ce livre, sa traduction et sa présentation, ainsi que toute autre suggestion.

Notre adresse:
Editions Mir,
2, Pervi Rijski péréoulok,
Moscou, I-110, GSP, U.R.S.S.

Imprimé en Union Soviétique